MODEL ENHANCEMENTS AND PARAMETER EXTRACTION FOR THE MMSPICE/QBBJT MODEL

BY

TZUNG-YIN LEE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1997

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my advisor, Dr. Robert M. Fox, for his devoted guidance, patient encouragement, and support throughout my Ph.D study. It is a great pleasure to work with him. I would also like to give thanks to my supervisory committee, Dr. Gijs Bosman, Dr. Jerry G. Fossum, Dr. Mark Law, and Dr. Mang Tia for their willingness to provide me their time and guidance.

I am grateful to the Semiconductor Resource Company (SRC), Motorola, and Texas Instruments, Inc. for their technical and financial support throughout my research. I feel especially obligated to express my deep appreciation to Dr. Mark Foisy at Motorola for his invaluable comments and discussions. I am also grateful to Mr. Tom Vrotsos, Dr. Keith Green and Dr. Zhiliang Chen, for their guidance during my internship in Texas Instrument Inc. Without their support, the work could never have been finished. I would also like to thank Dr. Ranjit Gharpurey for the invaluable discussion on BJT substrate resistance calculation.

I would like to recognize in this achievement Dr. Ming-Chang Liang and Messrs. David Zwidinger, Jonathan Brodsky, Ming-Yeh Chuang, and Amed Nadeem, who helped me in many ways with sincere and profound discussions on related topics, and my best friend, Everett Yang, who helped me in editing the drafts of this dissertation.

My gratitude is also extended to many of my friends, who have made my life most cheerful. I cannot mention them all, but I would like to mention Dr. Paul Chen, Dr. Simon

Wang, Dr. Bruce Liu, Jenfeng Yue, Jenyi Yue, and Chihou Vong. Without them, my life would have been difficult and barren.

My deepest gratitude goes to my parents Chen-I and Kuolan Lee. They presented themselves as an example in several ways that I can be humble and responsible. In addition, without their unfailing love and financial support, I could never have come to the United States and begun my graduate studies.

I am pleased that my wife Fiona has been able to share the achievement with me. Being a wife of a student for many years is not easy at all. Finally I would like to give my praises to my Lord, Jesus Christ, for He is the beginning and He is the end.

TABLE OF CONTENTS

page

ACKNOWLEDGEMENTS ii

ABSTRACT vi

CHAPTERS

1 INTRODUCTION 1

2 BACKGROUND, PHYSICS, AND MODEL FORMULATION OF THE MMSPICE CHARGE-BASED BIPOLAR TRANSISTOR MODEL - QBBJT ... 10

2.1 Background Review 10
2.1.1 Ebers-Moll Model 10
2.1.2 Gummel-Poon Model 11
2.2 Novel Compact Models for Bipolar Transistors 12
2.2.1 The Extended Gummel-Poon Model 12
2.2.2 MEXTRAM 13
2.2.3 MMSPICE/QBBJT Model 14
2.2.4 Vertical Bipolar Inter Company Model (VBIC95) 16
2.3 Model Formulation of QBBJT 17
2.3.1 BJT Operation Modes 17
2.3.2 Formulation of DC Current-Voltage Equations 22
2.4 QBBJT's Small-Signal Equivalent Circuit 24

3 MODEL ENHANCEMENTS OF THE QBBJT MODEL 30

3.1 Self-heating Modeling 30
3.2 DC Current Crowding Modeling 35
3.3 AC Current Crowding Modeling 44
3.3.1 Simple Model Formulation 44
3.3.2 Accounting for DC Current Crowding in AC Current Crowding Analysis 49
3.3.3 Model Implementation and Verification 52
3.3. 4 Conclusion and Discussion 55

4 PARAMETER EXTRACTION FOR MMSPICE/QBBJT 58

4.1 Introduction 58
4.2 Extraction of Gummel-Poon-Like Parameters 68
4.3 Extraction of Doping Profiles 70
4.3.1 Collector Doping Profile 70
4.3.2 Base Doping Profile 75
4.3.3 Integration of Doping Profile Extraction Procedures 76
4.3.4 Extraction Results and Discussion 77
4.4 Extraction of QBBJT's Small-Signal Equivalent Circuit 84
4.4.1 Methodology 84
4.4.2 Calibration of QBBJT's Extrinsic Resistance Parameters 88
4.4.3 Evaluation of QBBJT's Charge Partition Factors 92
4.5 Conclusion 95

5 SPICE MODELING OF BJT SUBSTRATE RESISTANCE 101

5.1 Introduction 101
5.2 Algorithm to Calculate BJT Substrate Resistance 104
5.3 Limitations of the Algorithm 111
5.4 AC Verification of the Algorithm 113
5.5 Including BJT Substrate Resistance in Parameter Extraction 116
5.6 Conclusions 118

6 SUMMARY AND SUGGESTED FUTURE WORK 119

6.1 Summary 119
6.2 Suggested Future Work 122

APPENDIX THE SUBSTRATE RESISTANCE CALCULATION PROGRAM SUBTEST 125

REFERENCES 128

BIOGRAPHICAL SKETCH 133

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MODEL ENHANCEMENTS AND PARAMETER EXTRACTION FOR
THE MMSPICE/QBBJT MODEL

By

Tzung-Yin Lee

May 1997

Chairman: Robert M. Fox
Major Department: Electrical and Computer Engineering

This dissertation addresses several enhancements for the MMSPICE/QBBJT model which are critical to contemporary bipolar transistors. First, self-heating is included by adding an extra thermal node into QBBJT's admittance matrix. Through the extra thermal node, the transistor's local temperature is simulated along with its terminal currents and voltages. Second, QBBJT's dc current crowding analysis is modified to better model the transistor's high injection behavior. Third, a compact and physical ac current crowding model is developed. AC current crowding is predominantly a nonquasi-static (NQS) behavior caused by the majority carriers transporting laterally along the base. By adapting the RC transmission line model, an equivalent circuit is developed to account for the NQS carrier transport. The equivalent circuit also accounts for the base resistance reduction due to dc current crowding. With the effective base width calculated from dc

current crowding analysis, the model can also be extended to account for the nonuniformity in the base-emitter admittance along the base.

This dissertation presents a complete parameter extraction methodology for the MMSPICE/QBBJT model. QBBJT is designed to take the model parameters derived from transistor's structural information such as layout and doping profile. In such a way, QBBJT enables simulation even when the process is not yet finalized. This attribute makes QBBJT especially useful in concurrent engineering applications. However, QBBJT still needs a methodology to calibrate or tune its parameters when representative measurements are available. This dissertation demonstrates a systematic methodology to extract QBBJT's model parameters from measurements. The methodology can be implemented in automated extraction tools. With this methodology, QBBJT now can be used in either the predictive mode for concurrent engineering applications or the precision mode for routine circuit verification.

Finally, the dissertation presents a systematic methodology to include substrate resistance in BJT parameter extraction and simulation. Substrate resistance is often neglected in compact BJT models, since it is layout dependent. Even for the most sophisticated parasitics extraction programs, substrate resistance is ignored because its calculation involves computation-intensive three-dimensional simulations. Based on a fast calculation algorithm which was initially developed to characterize substrate noise coupling in mixed-signal applications, the dissertation demonstrates an efficient way to eliminate the uncertainties resulting from the layout dependence of substrate resistance in BJT parameter extraction and circuit simulation.

CHAPTER 1
INTRODUCTION

Since the bipolar junction transistor (BJT) was invented in 1947, it has been employed in applications of analog amplification and digital switching. Complementary-metal-oxide-semiconductor (CMOS) technology, because of its characteristics of low-power consumption and high-packing density, is preferred in fabrication of ultra large scale integrated (ULSI) circuits. However, because of its advantages in high-speed operations, bipolar technology is still employed in a wide variety of applications, ranging from high performance mainframe computers and data links, to analog and digital telecommunications and wireless communications [1]. As aggressive scaling and advanced process technology continue to improve the speed-power performance of CMOS, CMOS threatens to overwhelm all the advantages of conventional bipolar transistors and become the ubiquitous technology. Nevertheless, the process and lithography breakthroughs can also improve the speed performance of bipolar transistors. Because bipolar transistors have advantages in operation speed over CMOS technology, they are often used in high-performance digital and low-cost high-speed RF applications.

As the result of its high power dissipation compared with CMOS technology, bipolar technology is rarely used in very high-level integration. To reduce power dissipation with high levels of integration, contemporary bipolar technologies have been developed to maintain BJT's speed advantages while achieving higher integration level.

One approach is to merge high-performance bipolar devices into CMOS processes, and form Bipolar/CMOS (BiCMOS) processes. In such processes, bipolar transistors are employed only where the leverage is high enough to justify extra power consumption. By retaining the benefits of bipolar and CMOS technology, BiCMOS technology allows overall speed-power-density performance previously unattainable with either technology. BiCMOS gates now can reach a speed about twice that of CMOS gates with approximately equal power dissipation [2]. Availability of bipolar transistors along with CMOS devices also adds design flexibility in analog or mixed-signal applications. However, the manufacturing cost of BiCMOS ICs is considerably higher than that of CMOS ICs. In addition, insertion of bipolar processes into baseline CMOS processes usually constraints optimization of either bipolar or CMOS transistor performance [3].

Another direction for contemporary bipolar technology is to fabricate high-performance vertical pnp as well as npn bipolar transistors on the same chip. Complementary bipolar technology has long held the promise of achieving high-speed with lower power dissipation. Until recently, the bottleneck of complementary bipolar technology is to realize high-performance vertical pnp transistors with base-line npn bipolar processes [1]. Therefore, complementary bipolar technology completely is unlikely to replace vertical npn technology in the near future.

Currently the most advanced technology for realization of high-speed bipolar transistors is based on double-poly self-aligned technology. Double-poly self-aligned bipolar processes were developed over a decade ago, and rapidly became the favored technology for manufacturing high-performance bipolar transistors. In this process, the first-layer poly forms the base contact. The base contact poly serves as local interconnect

and as a diffusion source to form low-resistive extrinsic base. Like the base contact poly, the second-layer poly forms the emitter contact and serves as a diffusion source to form a shallow emitter junction. A typical structure for a double-poly self-aligned bipolar transistors is shown in Fig. 1-1.

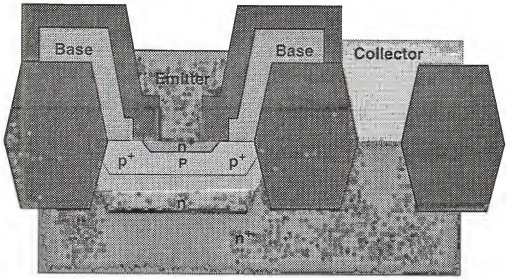


Figure 1-1 Cross-section of common advanced double-poly self-aligned bipolar transistors

Using highly doped poly-Si as an emitter contact reduces process complexity and maintains the separation between the metal contact and the very sensitive monocrystalline silicon region. The separation prevents sintering effects and metal spikes [4]. The poly-emitter-contact BJT also exhibits characteristics superior to its metal-contact counterpart. First, a shallow base junction can be formed with sufficient control on shallow emitter junction formation. The shallow base reduces base transit time and improves speed performance of the transistor. Secondly, bipolar transistors with poly emitter contacts

usually have better emitter efficiency than bipolar transistors with metal contacts, because of the lower carrier mobility in poly silicon and the interfacial oxide layer interposed between active emitter and poly emitter. As the thickness of emitter continues to decrease for shallower base junctions, conventional metal-contact transistors would suffer a serious current-gain reduction. This is because the emitter becomes transparent to the injected minority carriers from the base, when its thickness is less than the characteristic diffusion length of minority carriers. Finally, the poly emitter serves as a gettering source for metal impurities [4].

With double-poly self-aligned technology, BJT's lateral dimensions are also significantly reduced. This improves BJT's speed performance in several aspects. First, emitter width is narrower than the minimum line-width given by lithography. The reduction in emitter width reduces the intrinsic base resistance, thus easing the dc current crowding effect for BJT operating at high current densities. Secondly, the distance between the edges of emitter and base contacts, which is usually determined by the width of a spacer, is reduced so that BJT's extrinsic base resistance is reduced. Finally, as the extrinsic base area is reduced, the overall base-collector junction capacitance is reduced. The reduction in both extrinsic base resistance and base-collector junction capacitance serve to improve speed performance and power gain.

Reducing base thickness, which decreases base transit time, can further enhance speed performance. However, reducing base thickness always conflicts with the need to minimize intrinsic base resistance, which is essential in optimizing the BJT's power gain. In addition, reducing base thickness also reduces the voltage at which transistors break down because of base punch-through. A possible way to compromise speed, power gain,

and breakdown voltage in designing a BJT is to increase the transistor's base doping. However, increasing base doping sacrifices transistor's current gain.

Using SiGe epitaxial film for the transistor base in silicon-based process is a cost-effective solution to the above dilemma. Because the ratio between transistor's base and emitter current injections is exponentially amplified by the bandgap difference at the emitter-base junction, a SiGe heterojunction bipolar transistor (HBT) can have 10 to 100 times the current gain of a homojunction silicon bipolar transistor. This extra current gain can be traded for the speed performance by optimizing base doping and thickness. In addition, as the bandgap of SiGe can be manipulated by varying the Ge concentration in the Si base, a built-in electric field can be created to assist minority transport in base, which further enhance transistor's speed performance. Recently, SiGe HBTs with 75 GHz unity gain frequency have been reported.

Even for purely silicon technology, aggressively scaled double-poly self-aligned bipolar transistors now can reach unity-gain frequencies as high as 30 GHz [5], thus being suitable for very high-speed digital and analog applications. A decade ago these high-speed applications were reserved for only compound semiconductor devices. However, the device models, which are essential to simulation, design and optimization of bipolar devices and circuits, have not kept pace with the tremendous progress achieved in process technology.

In designing Bipolar/BiCMOS devices and circuits, technology CAD (TCAD), which explores and evaluates various design trade-off without time-consuming and costly device fabrication, is widely used in today's state-of-the-art design process. In the past, technology design was mostly based on experiment-based (i.e. trial-and-error)

optimizations. However, as the costs and duration of development increase with the complexities of today's Bipolar/BiCMOS technology, such a process is slow and expensive. TCAD predicts and optimizes the electrical performance of VLSI circuits without using expensive fabrications. TCAD reduces development costs, shortens development times, and helps maintain process control in design and production environments.

In conventional TCAD strategy, the output of a process simulator is applied as input to a device simulator, whose output in turn is used in extracting device model parameters for circuit simulations. The computational cost of the device simulation in this approach is very high, especially considering the inevitable uncertainty in doping profiles, even when they are tuned using SIMS other measurements.

MMSPICE developed in University of Florida is a mixed-mode device/circuit simulator to facilitate TCAD design processes. MMSPICE merges device physics into circuit simulations, therefore reducing the computation cost. Its charge-based bipolar transistor model QBBJT is based on regional analysis of ambipolar transport equations, subject to moving boundary conditions. The model is thus consistent with device physics implied by the drift-diffusion relations used in most device simulators, but runs much faster. QBBJT's model parameters are mostly derived from device's layout and doping profile. Therefore, correlations between model parameters and process information are automatically maintained. However, the use of MMSPICE/QBBJT model has been rather limited. This is because that the MMSPICE/QBBJT model lacks a well-defined parameter extraction methodology. In practice, the MMSPICE/QBBJT model is used in a variety of modes. In concurrent engineering applications, QBBJT takes parameter derived from

simulated doping profiles and runs at predictive mode. In this way, QBBJT enables circuit design before the process is finalized; the design cycle is significantly shortened. The ability to predict performance from device structure is very useful in the early stages of product design. However, once the technology is mature and test devices are available, simulation accuracy and efficiency become the key issues.

The dissertation presents a parameter extraction methodology that allows QBBJT model parameters to be calibrated in accordance with measured data, enabling accurate simulations even when the process information is incomplete or inaccurate. This ability to use model parameters derived either from process/structural information or from electrical measurements makes the QBBJT model uniquely flexible. Now a single compact model can be used in either a TCAD or best-fit-to-data mode. As the technology matures, only simple revisions to the process-derived parameters are need for transparent transitions between modes.

The dissertation also addresses several refinements to the MMSPICE/QBBJT model. First, self-heating is implemented in the QBBJT model by adding an extra thermal node to QBBJT's admittance matrix, which is used in SPICE's modified nodal analysis. Self-heating is inherent to all semiconductor devices. Self-heating is a significant effect even for bulk-silicon bipolar transistors operating at modest power dissipation. Ignoring self-heating in the BJT model can cause problems in simulating analog circuits because BJT's output conductance is either overestimated for homojunction silicon bipolar transistors or underestimated for some compound heterojunction bipolar transistors.

DC current crowding is an important two-dimensional effect. The original QBBJT model account for current crowding around the emitter periphery by using the effective

base-emitter voltage, which can significantly underestimate BJT's high injection effects. In this dissertation, we present an alternative way to model current crowding by using an effective emitter area. The approach is consistent with the original dc current crowding analysis, but maintains correct high injection analysis.

AC current crowding effect is a relatively new topic in compact bipolar transistor modeling. While dc current crowding has been studied for over 30 years, ac current crowding has received relatively little attention. AC current crowding is a nonquasi-static (NQS) effect, and is often overlooked in the BJT's small-signal equivalent circuit. This is because BJT's small-signal equivalent circuits are usually derived using quasi-static (QS) approximations. As the total base charge storage is modeled by a lumped capacitor, the time required for the majority charges in the base to redistribute laterally along the base is overlooked. The QS approximations is equivalent to assuming that the carriers can travel with infinite velocity in certain regions. This dissertation presents a new compact model to account the NQS ac current crowding. This model is consistent with QBBJT's dc current crowding analysis, and is easily realized in most compact bipolar transistor model.

As bipolar transistors continues to be scaled down for the demand of faster operation and higher integration, substrate resistance becomes more and more influential to the overall performance of bipolar transistors. For most compact BJT models, substrate resistance is not included by default. A common approach to model substrate resistance is to optimize the collector-to-substrate capacitance to partially account for its effect. However, the approach can cause numerous difficulties in statistical modeling and circuit design, because BJT's substrate resistance is layout dependent. In this dissertation, a systematic methodology to include substrate resistance in SPICE BJT simulation is

presented. With the methodology, BJT's substrate resistance is consistently included in BJT's parameter extraction and circuit simulations.

The dissertation is organized as follows: Chapter 2 reviews several classic and novel compact models. At the end of Chapter 2, QBBJT's model formulation is briefly discussed to introduce several concepts in the following chapters. Chapter 3 addresses several additions and refinements of QBBJT. Chapter 4 presents a complete parameter methodology for QBBJT. Chapter 5 presents a systematic methodology to include substrate resistance in BJT parameter extraction and circuit simulation. Chapter 6 concludes the dissertation with a summary and suggested future work.

CHAPTER 2
BACKGROUND, PHYSICS, AND MODEL FORMULATION OF THE MMSPICE CHARGE-BASED BIPOLAR TRANSISTOR MODEL - QBBJT

This chapter discusses several classic bipolar transistor models and a number of novel models designed to accommodate contemporary technology breakthroughs. This chapter provides concise comparisons between these models. The model formulation of the MMSPICE/QBBJT model is briefly reviewed at the end of this chapter to introduce several concepts elaborated in following chapters.

2.1 Background Review

2.1.1 Ebers-Moll Model

The Ebers-Moll model was an empirical model invented in early 50s [6]. The original Ebers-Moll model is only a dc large signal model which contains two diodes connected back-to-back with a controlled current source in parallel with each of them. The controlled current sources represent current diffusing from emitter to collector in forward operation and from collector to emitter in reverse operation. The expanded Ebers-Moll model has extra elements added to its equivalent circuit to account for resistive and capacitive parasitics. The addition of fixed terminal resistors allows the Ebers-Moll model the ability to account for nonideal ohmic voltage drop across resistive parasitics. By including several lumped capacitors, the expanded Ebers-Moll model can account for

charge storage in the quasi-neutral base and depletion regions. This enables ac simulations of bipolar transistors. With the addition of several empirical parameters, the expanded Ebers-Moll model accounts for several secondary effects. For instance, the expanded Ebers-Moll model uses forward and reverse Early voltages to account for base-width modulation and includes temperature coefficients for several parameters [7]. However, the Ebers-Moll model cannot sufficiently account for high current effects.

2.1.2 Gummel-Poon Model

As a physical supplement to Ebers-Moll model, the Gummel-Poon model was formulated based on the one-dimensional integral charge-control relation (ICCR) [8]. Many of the Gummel-Poon model's parameters are bias dependent. The Gummel-Poon model includes several empirical parameters, such as the knee current I_K, to account for high injection effects. Even though the Gummel-Poon model is the most popular model employed in a wide variety of circuit simulators, it has remained mostly unchanged for 20 years and is insufficient to model the advanced bipolar transistors.

The conventional Gummel-Poon model has the following major deficiencies: 1) It models base-width modulation (Early) effects using constant Early voltages (VAF and VAR), which results in unacceptable errors in simulating output conductances for BJTs with thin bases. 2) It uses constant collector resistance, thereby overlooking collector modulation and quasi-saturation effects. 4) Its small-signal equivalent circuit is derived based on quasi-static (QS) principles. It thus fails to account for nonquasi-static (NQS) effects in transistors operating at very high frequencies. 5) It fails to account for some

extrinsic parasitics, which are negligible in large devices, but important for scaled devices, e.g., the base-emitter side-wall capacitances. 6) It is not scalable with geometry.

2.2 Novel Compact Models for Bipolar Transistors

A predictive model for scaled bipolar transistors must account for (1) minority carrier transport in quasi-neutral base region for typical base doping profile and all injection levels; (2) majority carrier transport in quasi-neutral collector accounting for nonlinear voltage drops across the epi-collector and for quasi-saturation; (3) forward and reverse base-width modulation (Early) effects; (4) base widening (Kirk) effects; (5) impact ionization with consideration of nonlocal carrier energy transport relations; (6) two-dimensional geometric effects, such as side-wall current injection and current-crowding effects; (7) physical and continuously differentiable ac differential resistances and capacitances [9]; and (8) global and local thermal effects, i.e., self-heating. Many compact physical models based on these requirements have been proposed in the past ten years. The extended Gummel-Poon model [10], VBIC95 [11], MEXTRAM [12] and the MMSPICE charge-based bipolar transistor model (QBBJT) [14] are reviewed in the following sections.

2.2.1 The Extended Gummel-Poon Model

Using Kull's majority carrier transport equation in the quasi-neutral collector [10], the extended Gummel-Poon model accounts for quasi-saturation effect by calculating nonlinear voltage drop across the quasi-neutral collector. The extended Gummel-Poon model is based on the formulation of the original Gummel-Poon model and inherits most

of its parameters. The extended Gummel-Poon model is easily implemented and is familiar to most circuit designers. Consequently, the extended Gummel-Poon model has been included in a wide variety of circuit simulators.

To include Kull's collector transport equation in the extended Gummel-Poon model, one artificial node is introduced at the base-collector junction. While the original integral charge-controlled relation (ICCR) is employed to simulate the minority carrier transport in the quasi-neutral base, Kull's collector current equation is applied in the quasi-neutral collector. The collector current is then solved by forcing current continuity at the artificial node. The extended Gummel-Poon model models quasi-saturation as collector resistance modulation, thus reducing the empiricism of using constant collector resistance. However, the extended Gummel-Poon model cannot properly simulate the I-V characteristics of all possible operating regimes, because it assumes that the entire epi-collector is quasi-neutral. For example, the extended Gummel-Poon model cannot physically simulate current-voltage characteristics of BJT operating in the nonohmic quasi-saturation regime where an excess-carrier induced SCR forms near the highly doped buried layer due to carrier velocity saturation at high collector current density.

2.2.2 MEXTRAM

MEXTRAM model was developed by de Graaff and Kloosterman in Phillips Research Lab. MEXTRAM differs from the original Gummel-Poon model in a number of aspects [12]: (1) MEXTRAM does not use the integral charge control relation (ICCR). It models minority carrier transport and charge storage for an exponentially doped base, which is common in advanced bipolar transistors. (2) The terminal currents and charge

storages of MEXTRAM model are formulated as functions of the doping profile and bias condition. (3) Base-width modulation (i.e., Early effect) is not modeled with the constant Early voltages, but with the base-emitter and base-collector junction depletion charges.

The latest version of the MEXTRAM model also adds a unified majority carrier transport equation in the collector accounting for various conditions of quasi-saturation [13]. The MEXTRAM model, used in the circuit simulator Pstar (Phillips) and Microwave Design System (Hewlett-Packard), has many advantages over the Gummel-Poon model for simulating BJT's quasi-saturation and high current characteristics.

2.2.3 MMSPICE/QBBJT Model

The MMSPICE/QBBJT model was developed by Jeong and Fossum [14] at University of Florida. While QBBJT was being first implemented in SPICE2G.6, QBBJT is now available in at least one commercial SPICE simulator. QBBJT is similar to MEXTRAM in many aspects. Both QBBJT and MEXTRAM use de Graaff's equation to model minority carrier transport in the quasi-neutral base [15] and Kull's equation to model majority carrier transport in the quasi-neutral collector [10]. QBBJT, however, has several unique features. QBBJT uses moving boundary conditions for different bias conditions, thus including many physical effects such as the base-width modulation (Early) effects and the base widening Early (Kirk) effect. QBBJT also allows for possible SCR formation in epi-collector, thus eliminating the nonphysical assumption used in the extended Gummel-Poon model. QBBJT physically simulates BJT's current-voltage characteristics for all possible operating regions.

Most of QBBJT's parameters are derived from device structure (layouts) and process information (doping profiles). This makes QBBJT suitable for TCAD and concurrent engineering applications, where a simple pre-processor (SUMM) is used to generate QBBJT's parameters directly from the process simulator output or from measured doping profiles. QBBJT predicts circuit performance even in the absence of fabricated devices. Therefore, QBBJT can be used in concurrent engineering applications, which efficiently shortens production cycles. With parameters derived from the estimation of process information, circuit designs can begin before representative test chips are available.

QBBJT is also suited for statistical circuit design applications. Since most of QBBJT's model parameters come directly from the device structure and process information, statistical correlations between circuit performance and process variations are automatically maintained. Such correlations are not usually maintained in empirical models such as Gummel-Poon. MMSPICE is much more computationally efficient than device simulators. This makes accurate statistical analyses, that require many simulations (such as Monte Carlo), practical.

QBBJT is charge-based. Instead of using lumped capacitive elements and transit time constants, regional charge storages are formulated in accordance with bias conditions and device structural information. The regional charge storage provides non-reciprocal transcapacitances.

The QBBJT model was made more realistic and predictive for simulating advanced scaled BJTs by including nonlocal modeling of impact ionization and nonideal two-dimensional effects [16]. High electric fields and field gradients are common in BJTs

with aggressively scaled vertical dimensions. As electric fields change rapidly over distances comparable to the energy-relaxation mean-free-path, nonlocal effects such as carrier velocity overshoot have been shown to influence characteristics of BJTs significantly. Current crowding effects, caused by laterally distributed base resistance, is a prominent two-dimensional effect. Current crowding degrades BJT current gain and unity-gain frequency.

The QBBJT model was extended for SiGe HBT simulations by including bandgap variation parameters [17]. With the operation defined physically for the entire operating range in a unified and systematic way, the numerical robustness of QBBJT has been greatly enhanced. However, since QBBJT uses different model equations for different operating regions, first or higher order derivatives of currents and regional charges can still be discontinuous with improper model parameters. These discontinuities can cause numerical problems and errors in simulating analog circuits.

2.2.4 Vertical Bipolar Inter Company Model (VBIC95)

VBIC95 was proposed in 1995 by a group of IC and CAD industrial representatives as a replacement for the conventional Gummel-Poon model. The group evaluated compact BJT models developed since Gummel-Poon model. Their objective was to pinpoint the best model and bring it to the mainstream of bipolar IC designs. However, no models they evaluated satisfied all the requirements from the committee, so they decided to develop their own.

VBIC's formulation is similar to the extended Gummel-Poon model, although, with several unique features: 1) VBIC uses base-emitter and base-collector junction

depletion charges to model base-width modulation effect, thereby eliminating the empiricism of using constant Early voltages. 2) VBIC adds a parasitic pnp modeled with a simplified Gummel-Poon model. 3) To avoid numerical problems, VBIC uses single-piece, smooth functions whenever possible to model BJT's currents and small-signal capacitances. 4) VBIC includes self-heating.

Table 2-1 summarizes the above discussion contrasting different models. Notice that although some attributes can be included in almost all models, some attributes are pertinent to only specific types of models, i.e., empirical or physical models. This is because that they are just formulated differently. For example, an empirical model is usually continuously differentiable. However, its parameters are usually hardly scalable. Also, a physical model can include as many physical effects as possible. However, it is usually more computationally intensive than an empirical model.

2.3 Model Formulation of QBBJT

2.3.1 BJT Operation Modes

Before QBBJT's model formulation is reviewed, all possible operation modes of a BJT must be defined. Depending on whether a space-charge-region (SCR) is formed at base-collector metallurgical junction, BJT operation is divided into two distinct regions: saturation and forward regions. Different saturation conditions can be defined based on the velocity of majority carriers in epi-collector. When the majority carriers in epi-collector move with a velocity proportional to the local electrical field, the BJT is operating in ohmic quasi-saturation region and the epi-collector acts like a resistor. When the majority carriers reach their saturation velocity, excess majority carriers must be injected into the

Table 2-1 Comparison between compact BJT models

		Extended GP	MEXTREM	QBBJT	VBIC
Base analysis	Use Gummel-Poon ICCR	X			
	Use de Graaff's base transport equation		X	X	X
	Use Knee current to model injection	X			X
	Account for base push-out		X	X	
	Account for bandgap variation in base. SiGe HBT simulations.			X	
Early effect	Use Early voltages	X			
	Use depletion charges		X		X
	Use moving boundaries			X	
Collector analysis	Use Kull's equation to model quasi-saturation	X	X	X	X
	Account for various SCR formations in epi-collector		X	X	
	Account for field dependent mobility	X			X
AC analysis	Small-signal equivalent circuit derived with quasi-static (QS) approximation	X	X	X	X
	Bias dependent small-signal differential resistances and capacitances	X	X	X	X
	Account for vertical nonquasi-static (NQS) effects			X	X
	Account for lateral nonquasi-static (NQS) effects			X	

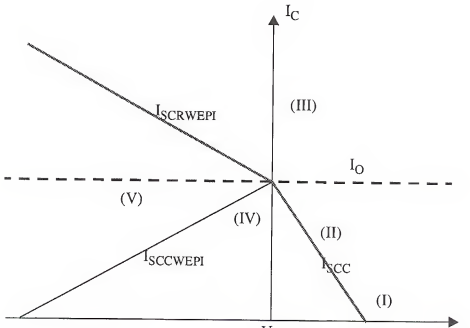
Table 2-1 -- continued

Other effects	DC current crowding			X	
	Transient current crowding			X	
	Self-heating			X	X
	Collector spreading effect		X		
	Two dimensional geometric (layout) effects			X	
General attributes	Continuous smooth terminal currents	X			X
	Continuous smooth differential resistances and resistances	X			X
	Scalable parameters			X	
	Numerical robustness	X			X

quasi-neutral collector, which induce a space-charged region (SCR) near the buried layer. In this situation, BJT is operating in nonohmic quasi-saturation region.

Different operating modes can also be defined for different bias conditions on an I_C-V_{BC} plane as shown in Fig. 2-1, where all the boundary lines are defined as [17]. Different regions in Fig. 2-1 are defined as following: (I) saturation; (II) ohmic quasi-saturation, where entire base and collector are quasi-neutral; (III) nonohmic quasi-saturation, where excessive charge is injected from base region and an SCR forms at the end of epi-collector; (IV) forward-active, where the base-collector junction is reverse biased and an SCR forms at base-collector metallurgical junction; (V) forward-active with the entire epi-collector depleted due to adequately high V_{CB}.

In Fig. 2-1, the thick solid boundary lines labeled ISCC and ISCCWEPI define the onset of SCR formation at base-collector junction. Above these two lines, no SCR is formed at the base-collector junction and BJT is operating at either ohmic or nonohmic quasi-saturation. Below these two lines, an SCR formed at base-collector junction, and the BJT enters forward-active operation. The thick dashed boundary line, $I_o = q \cdot A_C \cdot N_{EPI} \cdot v_{sat}$, is usually defined as the onset of base push-out. However, for collector currents greater than I_O, the quasi-neutral base does not necessarily extend into the epi-collector. Actually, base push-out depends on the collector current level as well as the base-collector bias voltage. I_O represents the maximum current that the epi-collector can carry before its majority carriers reach their saturation velocity, and above which extra carriers must be injected into quasi-neutral collector to sustain excess current.



(I): saturation region
(II): ohmic quasi-saturation region
(III): nonohmic quasi-saturation region
(IV): forward-active
(V): forward-active with the entire epi-collector depleted

Figure 2-1 Operation mode boundaries on I_C-V_{BC} plane.

2.3.2 Formulation of DC Current-Voltage Equations

The formulation of QBBJT's current-voltage equations and charge storages is based on de Graaff's equation to model minority carrier transport in the quasi-neutral base [12] and Kull's equation to model majority carrier transport in the quasi-neutral collector [10]. In deriving the minority carrier transport and charge storage equations, de Graaff assumed quasi-neutrality for an exponential base doping profile as

$$N_A = N_{AO} \cdot \exp(-\eta x / W_{BM}), \tag{2-1}$$

where N_{AO} is the extrapolated peak base doping, η is the base doping grading factor, and W_{BM} is the base metallurgical width. To account for intended or unintended bandgap variations which result in an aiding or retarding field in the quasi-neutral base, de Graaff's current equations are reformulated as

$$J_{nL} \equiv J_{nLF} - J_{nLR} \quad \text{for low injection conditions} \tag{2-2}$$

$$J_{nH} \equiv J_{nHF} - J_{nHR} \quad \text{for high injection conditions} \tag{2-3}$$

where

$$J_{nLF} \equiv qD_{nL}\eta' \frac{exp(\eta' W_b)}{exp(\eta' W_b) - 1} n(x_E) \tag{2-4}$$

$$J_{nLR} \equiv qD_{nL}\eta' \frac{n(x_C)}{exp(\eta' W_b) - 1} \tag{2-5}$$

$$J_{nHF} \equiv 2qD_{nH}\eta'' \frac{exp(\eta'' W_b)}{exp(\eta'' W_b) - 1} n(x_E) \tag{2-6}$$

$$J_{nHR} \equiv 2qD_{nH}\eta'' \frac{n(x_C)}{exp(\eta'' W_b) - 1} \tag{2-7}$$

and the base doping grading factor η' and η'' are defined as

$$\eta' = (\eta + \Delta E/kT)/W_{BM} \tag{2-8}$$

$$\eta'' = \Delta E/2kTW_{BM} \tag{2-9}$$

to account for bandgap variations. To ensure a smooth and continuous transition from low injection to high injection, linear extrapolation is used to get a single analytic current equation for all injection levels:

$$J_n = \frac{J_{nLF} \cdot N_A(x_E) + J_{nHF} \cdot n(x_E)}{N_A(x_E) + n(x_E)} - \frac{J_{nLR} \cdot N_A(x_C) + J_{nHR} \cdot n(x_C)}{N_A(x_C) + n(x_C)} \tag{2-10}$$

In the quasi-neutral collector, QBBJT uses an equation similar to Kull's formulation for the collector majority carrier transport equation:

$$I_C = \frac{qA\mu_n V_T N_{EPI}}{W_{QNR}}\left[K(\Psi_{BCO}) - K(\Psi_{BCI}) - ln\left[\frac{1+K(\Psi_{BCO})}{1+K(\Psi_{BCI})}\right] + \frac{\Psi_{BCO} - \Psi_{BCI}}{V_T}\right] \tag{2-11}$$

where

$$K(\Psi) \equiv \sqrt{1 + \left(\frac{2n_i}{N_{EPI}}\right)^2 exp\left(\frac{\Psi}{V_T}\right)} \ , \tag{2-12}$$

and W_{QNR} is the width of quasi-neutral collector.

QBBJT uses base-edge and collector-edge quasi-Fermi level separations at the base-collector junction, Ψ_{BCJ} and Ψ_{BCI}, to link the base and collector transport analyses. QBBJT then solves the collector current by forcing current continuity at the base-collector junction. When Ψ_{BCI} is greater than base-collector junction barrier, Φ_C, the BJT operates in saturation or quasi-saturation modes and Ψ_{BCI}=Ψ_{BCJ}. No SCR exists between the base and collector. On the other hand, BJT operates in the forward-active mode, where Ψ_{BCI} is modeled to be equal to Φ_C and Ψ_{BCJ} is modeled as

$$\exp\left(\frac{q\Psi_{BCJ}}{kT}\right) = \exp\left(\frac{qV_{BC}}{kT}\right) + \left[\exp\left(\frac{q\Phi_C}{kT}\right) - \exp\left(\frac{qV_{BC}}{kT}\right)\right]\frac{I_C}{I_{CRIT}} \quad (2\text{-}13)$$

where I_{CRIT} denotes the critical boundary currents (I_{SCC} or $I_{SCRWEPI}$) separating the quasi-saturation and forward-active operations as shown in Fig. 2-1.

QBBJT solves for the widths of base-emitter and base-collector SCRs to provide moving boundary conditions and to account for base-width modulation (i.e. Early effect) in base analysis. The width of base-collector SCR is calculated based on the potential drop across the base-collector SCR (i.e., Ψ_{BCI}-Ψ_{BCJ}). While taking into consideration the background charge modulation induced by collector current injection, QBBJT calculates the width of base-collector SCR using the depletion approximation. The base-emitter SCR, which controls the reverse Early effect, is modeled for both forward and reverse bias. With the boundary conditions defined physically, the QBBJT model properly links base collector transport equations and ensures current continuity throughout all operating regimes.

2.4 QBBJT's Small-Signal Equivalent Circuit

Before constructing the small-signal equivalent circuit, charge storages must be formulated. The formulation of excess charge stored in the quasi-neutral base is based on de Graaff's work. With a linear interpolation for intermediate injection, excess charge storage in the quasi-neutral base is formulated as:

$$Q_{nB} = \frac{N_A(x_E)}{N_A(x_E) + n(x_E)} \cdot Q_{nB}^L + \frac{n(x_E)}{N_A(x_E) + n(x_E)} \cdot Q_{nB}^H \quad (2\text{-}14)$$

where Q_{nB}^L and Q_{nB}^H are excess charge storages for low and high injection cases.

Excess charge stored in the emitter is formulated using the emitter minority transit time TE, an empirical parameter defined as:

$$Q_E = TE \cdot AE \cdot J_{EO} \cdot [\exp(V_{BE}/V_T) - 1]. \tag{2-15}$$

Excess charge stored in the base-emitter depletion region is modeled as:

$$Q_{JE} = q \cdot A \cdot \int_0^{x_E} N_A(x)dx \tag{2-16}$$

where $N_A(x)$ is the exponentially graded base doping profile and x_E is the base-side depletion width of base-emitter junction, modeled for both forward and reverse bias.

Excess charge storage in the quasi-neutral collector is modeled according to the electric field distribution shown as Fig. 2-2. In saturation and quasi-saturation (case A of Fig. 2-2), charge stored in the unmodulated region is zero because the electric field gradient is zero. The excess charge stored in the collector is only the charge in the modulated region, which is expressed as

$$Q_{QNR} = \frac{q^2 A^2 \mu_n^2 V_T n_i^2}{I_C}[\exp(V_{BCO}) - \exp(V_{BCI})] - \frac{qAn_i^2}{N_{EPI}} W_{QNR} \tag{2-17}$$

In nonohmic quasi-saturation (case B of Fig. 2-2) region, the excess charge in the collector is formulated as the sum of charge stored in the modulated region as (2-17) and in SCR at the end of the epi-collector as

$$Q_{EPI(SCR)} = W_{SCR} \cdot (qN_{EPI} - I_C/\upsilon_S), \tag{2-18}$$

where W_{SCR} is calculated with consideration of the background charge modulation caused by collector current injection.

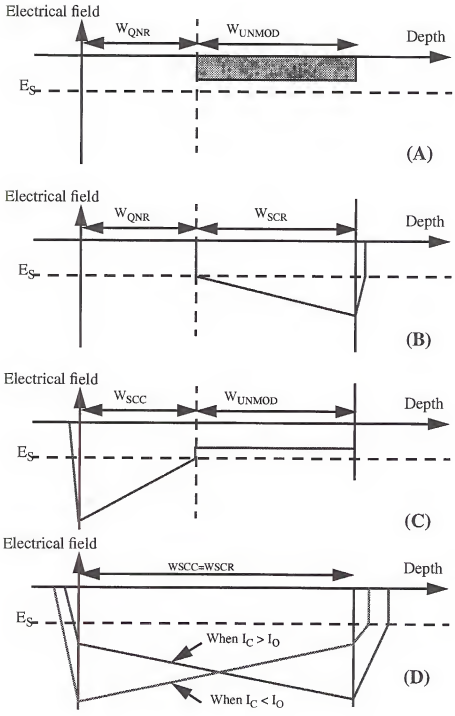


Figure 2-2 Electrical field distribution of each operation mode. A: saturation & ohmic quasi-saturation. B: nonohmic quasi-saturation. C: forward-active with epi-collector not fully depleted. D: forward-active with entire epi-collector depleted.

In forward-active mode (case C of Fig. 2-2), excess charge storage in the collector is formulated as charge stored in base-collector depletion region

$$Q_{EPI(SCC)} = W_{SCC} \cdot (qN_{EPI} - I_C/\upsilon_S). \tag{2-19}$$

where W_{SCC} is also calculated with consideration of background charge modulation caused by collector current injection.

In forward-active mode with the entire epi-collector depleted (case D of Fig. 2-2), in addition to charge stored in the epi-collector, which is modeled as (2-19) with W_{SCC} replaced by W_{EPI}, excess charge in the collector must include the charge stored in the buried layer as:

$$Q_{BL} = \frac{\varepsilon A}{W_{EPI}}(\Phi_C - V_{BC}) - \frac{W_{EPI}}{\upsilon_S}(I_O - I_C) - \varepsilon A E_S \tag{2-20}$$

While all regional charge storages are defined and formulated physically, next step to formulate the small-signal equivalent circuit is to partition these charge storages into different terminals. Charge partition is a typical technique to allocate a charge storage in a certain region into two different terminals so that the charge transport kinetics in the region can be partially accounted for. Through two empirical parameters ζ_B and ζ_C, Q_{nB} and Q_{QNR}, which represent the excess charge storage in the quasi-neutral base and in the extended base, are partitioned into the emitter and collector terminals.

The excess charge stored in the epi-collector Q_{EPI} is modulated by the collector current, which is a strong function of collector current. Q_{EPI} is allocated into the emitter and collector terminals as $Q_{EPI}^{E} = (W/\upsilon_s) \cdot I_C$ and $Q_{EPI}^{C} = -(W/\upsilon_s) \cdot I_O$, where W is the SCR width in the epi-collector.

The resulting charge storages in base-emitter and base-collector terminals are formulated as:

$$Q_{BE} = \zeta_B \cdot Q_{nB} + \zeta_C \cdot Q_{QNR} + Q_E + Q_{EPI}^E + Q_{JE} \tag{2-21}$$

$$Q_{BC} = (1-\zeta_B) \cdot Q_{nB} + (1-\zeta_C) \cdot Q_{QNR} + Q_{EPI}^C + Q_{BL} \tag{2-22}$$

where Q_{QNR} is set to zero for case (C) and (D). Q_{BL} is set to zero for case (A), (B) and (C).

With all terminal charges defined for all bias conditions, the small-signal equivalent circuit is formulated as shown in Fig. 2-3. This small-signal equivalent circuit has several unique features: (1) Implementation of the small-signal equivalent circuit is easy, since all the regional charge storages are derived from the dc solution; (2) The differential conductance (i.e., *dI/dV* terms) and capacitances (i.e., *dQ/dV* terms) are bias-dependent; (3) The transcapacitances are included explicitly; (4) The extra delay (phase shift) caused by base widening is included implicitly with Q_{QNR} modeled in quasi-neutral collector; (5) The NQS carrier transport inside the quasi-neutral base is taken into consideration through charge partitioning.

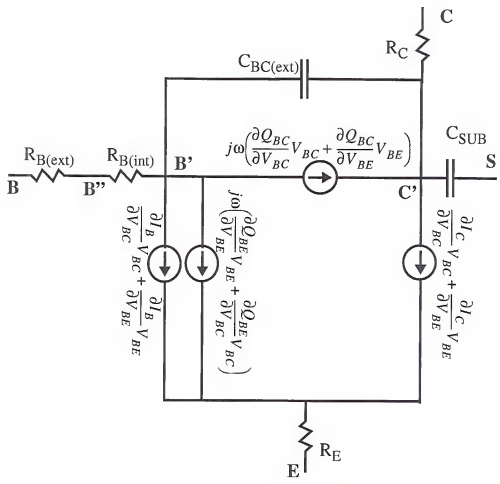


Figure 2-3 QBBJT's small-signal equivalent circuit

CHAPTER 3
MODEL ENHANCEMENTS OF THE QBBJT MODEL

3.1 Self-heating Modeling

This chapter discusses several additions and refinements to the MMSPICE/QBBJT model. Self-heating occurs in all semiconductor devices and circuits. Self-heating is an increase of device local temperature caused by device's own power dissipation. As packing density and power dissipation of VLSI circuits continue to increase with aggressive scaling, self-heating can become a severe problem and must be accounted for in simulations and circuit designs. Self-heating is often observed in MOSFETs as a drain current reduction, because the channel carrier mobility is degraded by the increased lattice scattering [18]. While decreasing the drain current in MOSFETs, self-heating increases the base and collector currents in bipolar transistors by enhancing the minority carrier injection at the base-emitter junction. Self-heating increases BJT power dissipation, thus causing positive feedback in BJT operation.

Since the mobility decreases slowly with temperature, the drain current reduction in MOSFETs is not easily detected unless there is a large temperature change of tens of degrees. Self-heating is especially important for MOSFETs built on SOI wafers, because the thermal resistance of silicon dioxide is about two orders of magnitude greater than that of silicon substrate. However, since the intrinsic carrier concentration increases

exponentially with absolute temperature, significant base and collector current changes in bipolar devices can be observed a small change in temperature.

Self-heating affects analog circuits significantly by increasing BJT's output conductance. This can lead to a significant reduction in the voltage gain in high-gain amplifiers [19]. Circuits that depend on close matching of base-emitter voltages can be very sensitive to self-heating, because small differences in junction temperatures can cause significant mismatch in currents. Recent research demonstrates that neglecting self-heating can cause significant errors in parameter extraction [20]. Self-heating can cause more than 50% difference in extracted values of BJT knee current. Therefore, proper models should include self-heating.

While the current gain of homojunction BJTs has a positive temperature coefficient, the current gain of heterojunction BJTs (HBT) can have either positive or negative coefficient. The negative coefficient is a result of the bandgap difference at the base-emitter junction [21]. Physical modeling of self-heating depends on accurate calculation of the device local temperature. To simulate the local temperature, one thermal node representing the temperature inside the bipolar transistor must be added into the QBBJT model, making it a 5 node model as shown in Fig. 3-1. Thermal impedance is the response of device temperature to variations in device power. With the thermal impedance calculated theoretically as in [22] or evaluated as discussed in [23], the equivalent circuit of Fig. 3-1 can be used for dc, ac and transient analyses.

In Fig. 3-1, the device local temperature is treated as an electrical voltage, the power dissipation as a controlled current source, and then the thermal impedance as an electrical impedance. The controlled current source I_{PWR} is calculated as

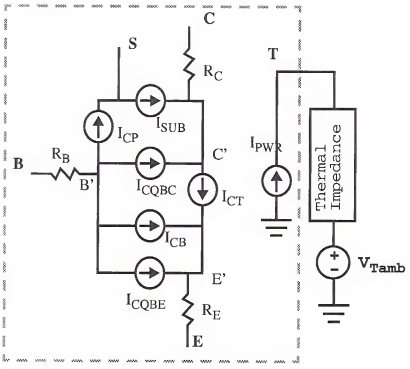


Figure 3-1 Schematic of thermal implementation

$I_{PWR} = I_C \cdot V_{CE} + I_B \cdot V_{BE}$. The local temperature is then evaluated through multiplying the power dissipation by the thermal impedance. In this way, self-heating simulation can be merged into the electrical simulation. In the implementation of self-heating, an extra node T was added into QBBJT by simply modifying QBBJT's admittance matrix in SPICE's modified nodal analysis. The implementation involves minor modifications in 18 subroutines and addition of a new subroutine to update temperature-dependent parameters in QBBJT model. Temperature dependent parameters are updated based on the temperature obtained from the thermal node T in every Newton-Raphson iteration. The resulting operating point is, thus, consistent with device's local temperature.

In this implementation, the ambient temerature is simulated as a voltage source V_{tamb}. In this way, we can simulate a BJT transistor at different ambient temperatures by simply varying the voltage source V_{Tamb}. This implementation allows easier extraction of the temperature dependence of BJT electrical characteristics than using the TEMP control card. For instance, we can vary the ambient temperature from -73 °C to 27 °C with a step of 5 °C using the DC sweep control as

.DC VTAMB -73 27 5

.PRINT DC I(VC)

In this way, we obtain directly the temperature dependence of the collector current. This also adds the flexibility to vary ambient temperature in frequency domain or time domain analyses. However, this implementation disables the original TEMP control card.

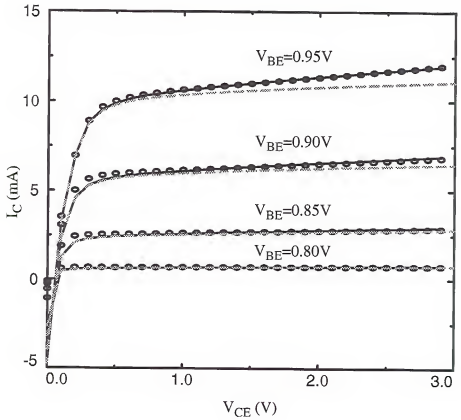


Figure 3-2 Output characteristic plot: Symbol lines denote measurement data for BJT of 1.6μmx20μm emitter from Motorola 0.5mm BiCMOS process. Solid lines are the MMSPICE simulation results with self-heating. Dashed lines are simulation without self-heating.

Fig. 3-2 shows a simulated output characteristic demonstrating self-heating in a BJT. The thermal resistance is calculated with the method in [22]. The thermal resistance can also be extracted using temperature dependent measurements [23].

3.2 DC Current Crowding Modeling

Current crowding caused by laterally distributed base sheet resistance is an important 2-D effect [24][25][26]. A half cross-section is shown in Fig. 3-3. As an ohmic drop results from dc base current flowing across intrinsic base, base-emitter voltage V_{BE} can be nonuniformly distributed. This thus causes the emitter current injection in the periphery of the base-emitter junction to be greater than that in the center. DC current crowding is especially severe at high current levels, because the transverse ohmic voltage drop is significantly higher at high current levels. Current crowding can be observed by measuring the low-frequency ac base resistance r_b, which decreases with dc emitter current I_E [24]. DC current crowding is usually associated with BJTs with large emitter widths. However, base thicknesses are scaled down to reduce base transit time for better speed performance, serious emitter current crowding occurs even for BJT's with moderate emitter widths.

Current crowding was first investigated by Pritchard in late 50's [24]. By cascading an infinite number of bipolar transistors with series base resistances, Pritchard derived the equivalent base resistance as one third of total base resistance. However, this result is invalid for the high base-current levels, which is typically involved in current crowding. Another analysis of emitter current crowding was done by Hauser [26]. Hauser manipulated the differential current-voltage relations across the intrinsic base region. With

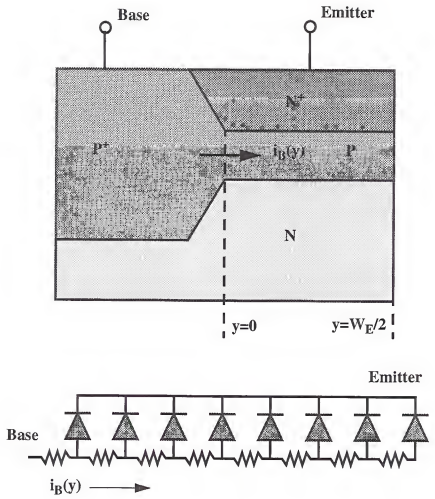


Figure 3-3 Half cross-section of a vertical bipolar transistors.

appropriate boundary conditions for the periphery and at the center, Hauser derived an analytical solution for total base current. Hauser's method is useful, but inadequate for advanced bipolar transistors, because it assumes that the base sheet resistance is constant, thus overlooking the base majority conductivity modulation, which occurs in high injection or base widening. QBBJT's dc current crowding model is mostly based on Hauser's approach. However, QBBJT includes consideration of base widening and majority carrier modulation. QBBJT calculates quasi-neutral base width and majority carrier concentration for each dc bias point. The bias-dependent base sheet resistance thus eliminates the inadequacy of Hauser's model.

Current crowding often enhances high injection effects [31], because the peak current density in the periphery is enhanced by current crowding. Emitter current crowding is usually modeled with an effective base resistance in SPICE and effective base-emitter voltage in MMSPICE. These approaches are equivalent to averaging the nonuniformly distributed emitter current injection along the base, thus underestimating the peak current injection in the periphery, as shown in Fig. 3-4.

To avoid the underestimation of the high injection effects, effective emitter area is used to model dc current crowding instead of effective V_{BE} or effective base resistance. Effective emitter width was first derived by Hauser [26]. With the consideration of conductivity modulation, the effective base width is derived similarly to [16] and [26]. This model is consistent with previous QBBJT dc current crowding analysis, but better simulates the current-crowding induced high-injection effects.

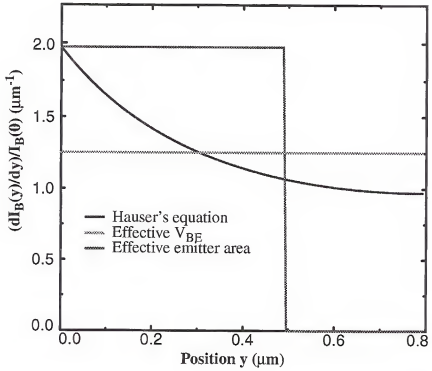


Figure 3-4 Two different models to approximate nonuniform current distribution along intrinsic base: (1) effective base-emitter voltage $V_{BE(eff)}$ model, (2) effective emitter area $A_{E(eff)}$ model.

In our current crowding analysis, the base-emitter junction is treated as an infinite number of diodes and resistors cascaded in series as shown in Fig. 3-3. Therefore, the base-emitter junction voltage as a function of position y, v(y), can be expressed as

$$v(y) = V_{BE} - \int_o^y i_B(\zeta) \cdot \rho d\zeta \ , \tag{3-1}$$

where ρ is the specific base resistivity and is expressed as

$$\rho = W_E / (2 \cdot \mu_p \cdot (Q_{BB} + Q_{QNR})), \tag{3-2}$$

Q_{BB} is the total charge of majority carrier in quasi-neutral base, and Q_{QNR} is the total charge of majority carrier in extended quasi-neutral base. Q_{BB} can Q_{QNR} are calculated from dc operating point, therefore making base resistivity bias-dependent.

From Fig. 3-3, the base current as a function of position y can be expressed as

$$i_B(y) = i_B(0) - \int_0^y 2 \cdot L_E \cdot J_{EO(eff)} \cdot \left[\exp\left(\frac{v(\zeta)}{V_T} \right) - 1 \right] d\zeta \, , \tag{3-3}$$

where L_E is the length of the emitter area and $J_{EO(eff)}$ is the effective saturation current density of diode, which is the steady-state diode saturation current plus the transient term obtained from previous time-step. Substituting (3-1) into (3-3) and differentiating the resulting equation twice, we obtain

$$\frac{d^2 i_B}{dy^2} + \frac{\rho}{V_T} i_B \cdot \frac{di_B}{dy} \, , \tag{3-4}$$

where we assume that $\exp(v(y)/V_T) \gg 1$. The general solution to (3-4) can be expressed as

$$i_B(y) = A \cdot \tan\left[\frac{A\rho}{2V_T} B \left(1 - \frac{y}{B} \right) \right] \tag{3-5}$$

where A and B are two unknowns. At $y=W_E/2$, $i_B(W_E/2)$ should decrease to zero. Therefore,

$$i_B(W_E/2) = 0 \tag{3-6}$$

Substituting (3-6) into (3-5), we can obtain the unknown B, and (3-5) becomes

$$i_B(y) = A \cdot \tan\left[z\left(1 - \frac{2y}{W_E}\right)\right] \tag{3-7}$$

where $z = (A\rho W_E)/V_T$. Hence, the total base current at y=0 can be expressed as

$$i_B(0) = A \cdot \tan(z) \tag{3-8}$$

To solve the last independent unknown A, we can substitute (3-7) into the first two coupled integral equations (3-1) and (3-3), and obtain

$$i_B(0) = L_E W_E J_{EO} \exp\left(\frac{V_{BE}}{V_T}\right) \frac{\sin(2z)}{2z}. \tag{3-9}$$

By equating (3-8) and (3-9), we can solve A and the dependent variable z, thereby obtaining the total base current for a given base-emitter voltage V_{BE}.

Notice from (3-9) that current crowding makes the total base current smaller than that expected from bipolar transistors with zero base spreading resistance. In the previous version of QBBJT, dc current crowding is modeled with an effective base-emitter voltage $V_{BE(eff)}$, that is

$$i_B(0) = L_E W_E J_{EO} \exp\left(\frac{V_{BE(eff)}}{V_T}\right), \tag{3-10}$$

where

$$\exp(V_{BE(eff)}) = \exp(V_{BE}) \frac{\sin(2z)}{2z}. \tag{3-11}$$

The result from the $V_{BE(eff)}$ approach is similar to the result using base resistance in SPICE. Both underestimate the peak current injection level. A more physical model of current crowding should evaluate effective emitter area, thereby maintaining current density increase in the periphery due to the reduction in effective emitter area as Fig. 3-4.

To account for current crowding effect using effective emitter area, we inherit most of the original model formulations. However, we use an effective emitter area $A_{E(eff)}$ instead of effective base-emitter voltage $V_{BE(eff)}$ in (3-10). The resulting effective emitter area $A_{E(eff)}$ is

$$A_{E(eff)} = A_E \frac{\sin(2z)}{2z} = W_E L_E \frac{\sin(2z)}{2z} \tag{3-12}$$

With A_E replaced by $A_{E(eff)}$ in the subroutine QBCT, QBBJT's current crowding model is modified for both DC and transient analysis. Fig. 3-5 and Fig. 3-6 show simulation results with the new current crowding modeling. The new model better simulates the current-crowding induced high-injection effect, i.e. high-injection roll-off, in current gain β and unity-current-gain frequency f_T.

In the original development of MMSPICE, the use of effective emitter area was thought of to cause instability. For this reason, transient simulations were performed on different devices to check numerical stability of the new current crowding model. The results show that the numerical stability of the new current crowding model is at least as good as that of the old one. For large devices (WE > 1.0 μm), the new current crowding model actually improves convergence stability.

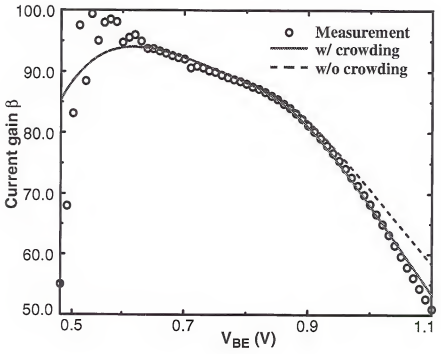


Figure 3-5 Current gain β vs. base-emitter voltage V_{BE}. From this figure we can see current-crowding induced high-injection effect is better manifested using the new current crowding model.

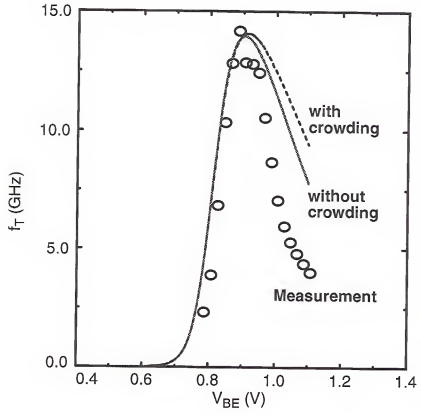


Figure 3-6 Unity-current-gain frequency f_T vs. base-emitter voltage V_{BE}. From this figure we can see current-crowding induced high-injection effect is better manifested using the new current crowding model.

3.3 AC Current Crowding Modeling

3.3.1 Simple Model Formulation

QBBJT's ac analysis does not include small-signal intrinsic base resistance, when dc current crowding analysis is active, which underestimates the total small-signal base impedance. To model small-signal base resistance, we must account for distributed base resistance as well as base-emitter capacitance (including both depletion and diffusion capacitances), which can lead to significant ac current crowding. Consider the distributed base-emitter RC network shown in Fig. 3-7. Such a network can be derived directly by linearizing the cascaded diode-resistor network as shown in Fig. 3-3. As the small-signal admittance of the base-emitter capacitance increases with frequency, small-signal base current tends to gather near the periphery as shown in Fig. 3-8. The phenomenon of nonuniform distribution of ac base current is called ac current crowding. As dc current also tends to crowd around the periphery of base-emitter junction (e.g. edge current crowding), the diffusion capacitance in the periphery is proportionally higher than that in the center. Therefore, ac current crowding is more severe with significant dc current crowding [27].

The phenomenon of dc current crowding was reported more than 30 years ago [26] and has been widely studied [27]. AC current crowding, however, has received much less attention than its dc counterpart. AC current crowding is a nonquasi-static (NQS) effect. The NQS effect is the time-dependent behavior cannot be obtained by extrapolating from steady-state analysis. In quasi-static (QS) charge-based modeling, different charging and discharging dynamics are modeled with transit time constants, which are calculated from a dc solution for excess charge storage. For instance, the base transit time can be calculated

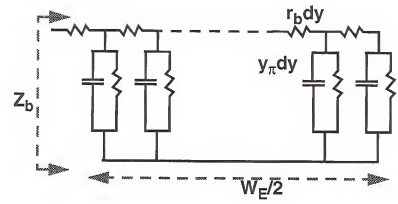


Figure 3-7 Small-signal equivalent circuit of the base resistance plus base-emitter capacitance.

*

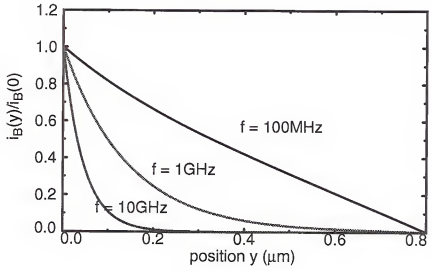


Figure 3-8 Normalized small-signal ac base current distribution along the intrinsic base. Calculated for typical BJT without significant dc current crowding.

as dQ_B/dI_C, where Q_B is the excess charge storage in the quasi-neutral base and I_C is the collector current. Actually, this base transit time is just the time required for the collector current to deliver charge to the quasi-neutral base. This approach thus neglects the time required for those charges to redistribute in the quasi-neural base.

The NQS charge redistribution time is especially important as the operating frequency approaches the reciprocal of carrier transit times. Since the lateral dimension of the base can be greater than its vertical dimension, the time required for the charge to redistribute laterally can be larger than the base transit time, which is predominantly controlled by the vertical dimension of the base. Therefore, the NQS effect caused by the lateral redistribution time can be significant, even when the BJT is operating at frequencies far below its unity-gain frequency f_T. As transistors are used at frequencies approaching their f_Ts, the lateral NQS effect must be carefully accounted for.

A common way to account for NQS effects is to use charge partitioning, which accounts for charge redistribution in a region by allocating the excess charge storage in a region to two different terminals. Charge partitioning is often used to account for NQS minority carrier transport in the quasi-neutral base, i.e. for vertical NQS effect [30]. Charge partitioning technique can also be adapted to account for majority charge redistribution in the quasi-neutral base, i.e. for the lateral NQS effects. In most compact BJT models, base-collector capacitance is partitioned into an intrinsic part and an extrinsic part, where the charge partition factor, i.e. XCJC in the Gummel-Poon model, is obtained from optimization of small-signal s-parameters. Actually, this is merely a way of using charge partitioning to account for the lateral NQS effect occurring in the intrinsic base to better fit the input small-signal parameters, i.e. s11 or s12.

To exactly represent the NQS charge redistribution time along the lateral base in ac current modeling would require cascading an infinite number of RC networks in series [28] [29]. Based on the series expansion of the input impedance of an RC transmission line, Versleijen derived a simple RC network, which consists of a resistor $R_b/3$ and capacitor $C_\pi/5$ in parallel, to simulate the ac current crowding effect [28]. This approach is simple and useful. However, this model is not sufficient at high base current, because it does not account for the differential base resistance reduction caused by dc current crowding. Also, it does not account for the non-uniformity in the base-emitter admittance caused by dc current crowding.

The new approach described below is based on [28]. However, by careful manipulation of the series expansion of the RC transmission line input impedance, the new model considers the differential base resistance reduction. Also, with the effective base length obtained in the dc current crowding analysis, the new model accounts for the non-uniform distribution of base-emitter admittance.

Under linear and small-signal conditions, the input impedance of a uniformly distributed transmission line can be expressed as

$$Z_b = \sqrt{\frac{R_b}{Y_\pi}} \coth \sqrt{Y_\pi R_b}, \tag{3-13}$$

where $Y_\pi = 1/R_\pi + j\omega C_\pi$ is the total base admittance, and R_b is the intrinsic base resistance. (3-13) is equivalent to the ac base impedance derived by Pritchard [24]. Using the series expansion of $\coth(x) = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \ldots$, (3-13) can be expressed as

$$Z_b \cong \frac{1}{R_\pi^{-1} + j\omega C_\pi} + \frac{R_b}{3} - \frac{R_b^2 \cdot (R_\pi^{-1} + j\omega C_\pi)}{45}. \tag{3-14}$$

Through simple algebraic manipulation, (3-14) can be further transformed as

$$Z_b \cong \frac{1}{R_\pi^{-1} + j\omega C_\pi} + \frac{1}{(R_b/3)^{-1} + (5R_\pi)^{-1} + j\omega(C_\pi/5)} \tag{3-15}$$

Then the simplified equivalent circuit of the distributed RC base impedance is shown in Fig. 3-9. The equivalent circuit of Fig. 3-9 is similar to that in [28], but contains an extra component of $5R_\pi$. At low current, the base-emitter resistance R_π is usually much larger than R_b so that this $5R_\pi$ term can be neglected. However, since R_π decreases with base current, the magnitude of R_π can become much smaller than R_b at high current. Therefore, $5R_\pi$ should be kept in the equivalent circuit.

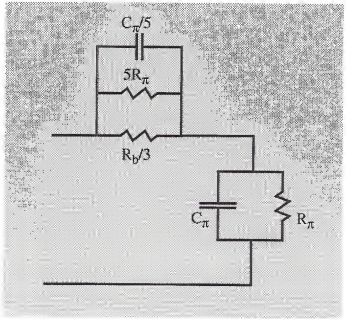


Figure 3-9 Equivalent circuit of (3-15).

The $5R_\pi$ term is to account for base-resistance reduction due to dc current crowding at high currents. In the equivalent circuit of Fig. 3-9, $R_b/3$ is the low-current dc intrinsic base resistance, which was first derived by Pritchard in 1958 [24]. However, at high currents, dc equivalent base resistance should be lowered, because the effective base width is reduced. By including $5R_\pi$, the simple equivalent circuit of Fig. 3-9 accounts for the dc base-resistance reduction due to dc current crowding at high currents.

The $C_\pi/5$ term is used to model the NQS ac current crowding. At low frequency, the base impedance of the equivalent circuit of Fig. 3-9 is merely the resistance of the two resistors ($5R_\pi$ and $R_b/3$) in parallel, because $C_\pi/5$ can be thought of as an open element. This implies that no significant ac current crowding occurs at low frequency. However, the effective base impedance becomes smaller at high frequency, because the ac current tends to crowd around the periphery, thus narrowing the effective base. With $C_\pi/5$ which shunts the effective dc base resistance ($5R_\pi$ and $R_b/3$ in parallel), the equivalent circuit of Fig. 3-9 properly simulates ac current crowding. However, the $C_\pi/5$ term should not be thought of as a physical component, but as an empirical component to approximately simulate the nonquasi-static (NQS) delay or phase shift.

3.3.2 Accounting for DC Current Crowding in AC Current Crowding Analysis

In the above discussion, the impedance of the base-emitter pn junction was assumed to be uniformly distributed; that is, the small-signal base-emitter resistance and capacitance, R_π and C_π, are constant along the lateral base dimension. This assumption is valid only for low currents, where V_{BE} is almost a constant along the lateral base; i.e.,

there is no significant dc current crowding. Since R_π and C_π are exponentially bias-dependent, they are nonuniformly distributed along the lateral base dimension, even for very small variations in V_{BE} due to dc current crowding. The net effect of the nonuniformly distributed base-emitter impedance on ac current crowding can be described as follows. Since the dc base-emitter voltage V_{BE} at the periphery is higher than that at the center, the small-signal base-emitter impedance is smaller than that in center, thereby causing more of the ac base current tends to flow through the periphery instead of the center. Hence, dc current crowding enhances ac crowding.

It is not easy to model the nonuniformly distributed base-emitter impedance, since, unlike the uniform transmission, no analytic solution exists. Nevertheless, little modification in the solution of uniform transmission line is needed to approximate the nonuniform transmission line for the purpose of compact modeling. For a physical approximation, we first examine the net effect of dc crowding on ac current conduction. The net effect of dc current crowding is to push ac current toward the periphery. This is equivalent to compress the area where most of the ac base current flows. Hence, we employ the concept of effective area, formulated for dc current crowding, to approximate the nonuniform transmission line. To accommodate the effective area factor in the previous ac current crowding analysis, we adjust the effective length of the RC transmission line used in our previous ac current crowding analysis.

To verify the ac current model, we compare the solutions of a nonuniform transmission line and a uniform transmission line with an effective length. The reason we use this comparison as a preliminary verification is that the solutions of transmission lines can better illustrate the concept of effective length than numerical device simulation, the

results of which are not always easily interpreted with regard to the intrinsic physics. We will use numerical device simulation as a final check to the model implementation.

To derive the differential equation which controls current-voltage relation of the nonuniform transmission line, the small-signal base-emitter admittance is modeled as

$$y_\pi(y) = y_{\pi o}\exp\left(\frac{V_{BE}(y)}{nV_T}\right), \tag{3-16}$$

where $y_\pi = 1/R_\pi + j\omega C_\pi$, $y_{\pi o}$ is a constant, and n is the injection ideality factor. n equals to one at medium injection level. At high injection, n is between one and two. However, for simplicity, we use n equal to one for all the injection levels.

To obtain the base-emitter voltage as a function of position, we can substitute (3-7) into (3-1) and obtain

$$V_{BE}(y) = V_{BE}(0) - 2V_T\ln\left\{\frac{\cos[z(1-2y/W_E)]}{\cos(z)}\right\} \tag{3-17}$$

where z is defined in (3-7). Then using (3-17) in (3-16), the small-signal base-emitter impedance as a function of position is expressed as

$$y_\pi(y) = y_{\pi o}' \cdot \{\sec[z(1-2y/W_E)]\}^2 \tag{3-18}$$

where $y_{\pi o}' = y_{\pi o}\exp[V_{BE}(0)/V_T][\cos(z)]^2$.

From Fig. 3-7, we can derive a pair of first-order differential equations for the small-signal current-voltage relation along the lateral base as

$$\frac{di_b(y)}{dy} = -y_\pi(y)v(y) = -y_{\pi o}' \cdot \{\sec[z(1-2y/W_E)]\}^2 \cdot v_b(y) \tag{3-19}$$

$$\frac{dv_b(y)}{dy} = -r_b \cdot i_b(y) \tag{3-20}$$

Substituting (3-19) into (3-20), we can obtain

$$\frac{d^2}{dy^2} v_b(y) = r_b \cdot y_{\pi o}' \cdot \{\sec[z(1-2y/W_E)]\}^2 \cdot v_b(y). \qquad (3\text{-}21)$$

The above second order differential equation is similar to that of a uniform transmission line, but does not have an analytical solution because of the secant function. (3-21) can be solved numerically using the boundary value method. The boundary conditions of (3-21) are described as following: At y=0, v_b(y)=1, if a unity ac voltage source is applied. At y=W_E/2, i_b(y)=0.

Fig. 3-10 shows that the small-signal voltage distribution along a nonuniform transmission line can be modeled approximately by a uniform transmission line with an effective length calculated from the dc current crowding analysis. Also, as shown in Fig. 3-11, the input admittance of a nonuniform transmission line as a function of frequency can be simulated properly by using a uniform transmission line with an effective length.

3.3.3 Model Implementation and Verification

To include the ac current crowding model into MMSPICE/QBBJT model, we implemented the equivalent circuit of Fig. 3-9 in QBBJT's ac admittance matrix in SPICE's modified nodal analysis.

MEDICI was used to verify the ac current crowding model and its implementation in MMSPICE/QBBJT model. In spite of being, strictly speaking, a quasi-static (QS) simulator, MEDICI can be used to simulate the nonquasi-static (NQS) extra phase-shift due to the charge redistribution in the device. The reasons are described as following: 1) MEDICI's mesh is usually small enough that the time required for the charge to be

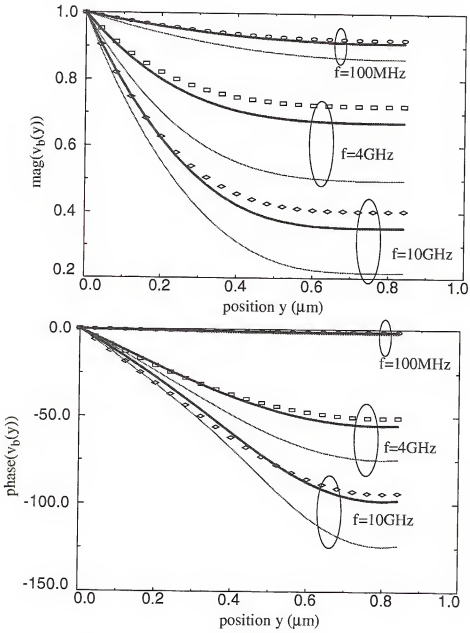


Figure 3-10 Comparison between the numerical solution of (3-21) and the analytical solution of uniform transmission line. Symbol lines represent the exact solution. Thick lines represent the analytic solution of the transmission line with effective length. Thin lines represent the analytic solution of the transmission line with normal length.

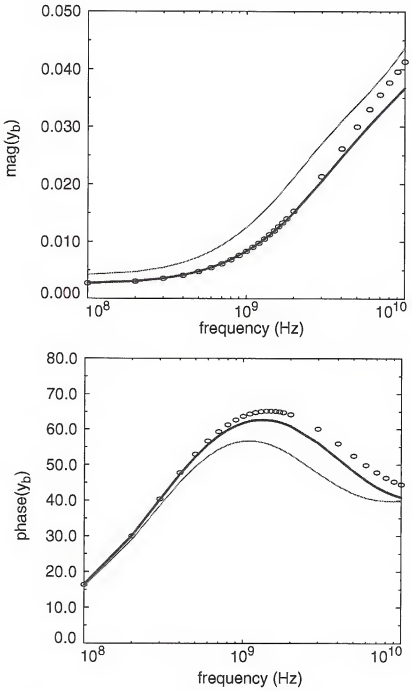


Figure 3-11 The input admittance of the nonuniform transmission line (symbol lines), the uniform transmission line with effective length (thick lines), and the uniform transmission line with normal length (thin lines).

redistributed in it can be assumed much smaller than any of device's characteristic transit time constants; therefore, we can assume that charges are instantaneously redistributed inside a mesh. 2) MEDICI's ac admittance matrix actually maintains all the interactions between adjacent meshes. Therefore, MEDICI enables simulations of NQS phase-shift due to the limited carrier transport velocity.

The MEDICI-simulated and MMSPICE-simulated base input impedances are shown in Fig. 3-12. While the magnitude of the base input admittance is only slightly influenced by ac current crowding, overlooking it causes significant errors in phase predictions at high frequency.

3.3.4 Conclusion and Discussion

This chapter describes addition of self-heating to the MMSPICE/QBBJT model. DC current crowding was modified to better simulate high injection effects. A compact equivalent circuit model was developed to model NQS ac current crowding effect. The model is consistent with the dc current crowding analysis and is easily implemented. Using the effective base width calculated from the dc current crowding analysis, the ac current crowding model can be extended to account for nonuniformly distributed base-emitter differential admittance along the lateral base. The resulting ac current crowding model is verified against numerical solutions of a nonuniformly distributed transmission line and MEDICI's simulations.

Notice that the NQS ac current crowding analysis may not be consistent with the existing transient crowding analysis. The ac and transient current crowding models emphasize different device physics for different applications. In fast switching (digital) applications, even when dc current crowding is not significant, the transient base current

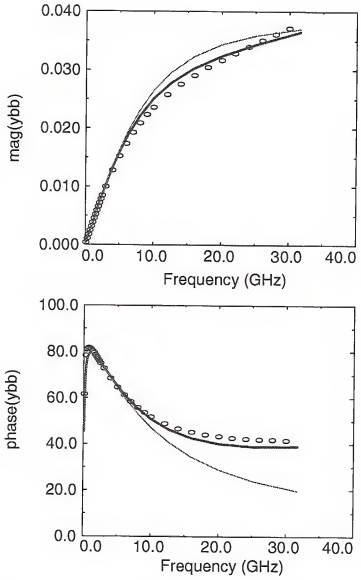


Figure 3-12 Comparison between the MEDICI simulated base input admittance and the MMSPICE simulated base input admittance with and without ac current crowding. The symbol line represents the MEDICI simulation. The thick solid line represents the MMSPICE simulation with ac current crowding. The thin line represents the MMSPICE simulation without ac current crowding.

can be very large causing severe transient crowding either in the periphery during turn-on or in the center during turn-off. Hence, in fast large-signal switching applications, the key need is to properly simulate the transient crowding. For simplicity, QBBJT calculates transient base current by differentiating majority charge storage based on the solution from the previous time step. This is equivalent to treating the whole quasi-neutral base as a single capacitor, ignoring the physics of charge redistribution along the lateral base dimension. In ac analysis, the ac current is assumed small enough that the excess charge distribution does not significantly deviate from its steady-state solution, so the primary issue is to account for the small amount of charge redistribution along the lateral base dimension. The focus of the NQS ac current crowding analysis discussed here is different from that of the transient crowding analysis. Therefore, the NQS ac current crowding analysis is not necessary to be consistent with the transient crowding analysis.

CHAPTER 4
PARAMETER EXTRACTION FOR MMSPICE/QBBJT

4.1 Introduction

QBBJT is implemented in at least one commercial SPICE version, and MMSPICE is available from its authors. QBBJT contains many features not provided by other available models, including accurate physical accounting for all quasi-saturation mechanisms, emitter current crowding, base push-out, and self-heating. In spite of these advantages, use of QBBJT has been rather limited. One probable reason is that QBBJT lacks a well-defined parameter-extraction methodology. In practice, even if QBBJT can be used with the model parameters derived from the device structures and process information, a circuit designer would still prefer a model with its model parameters calibrated for a given process to maximize the accuracy of circuit simulations. This chapter presents a systematic parameter extraction methodology for the QBBJT model parameters to be calibrated for a given process and used in its precision mode, even when the process information is not complete. The parameter extraction methodology, thus, successfully merges the QBBJT model into contemporary TCAD process and eliminates the need of involving another compact model with well-defined parameter methodology for precision purpose.

QBBJT's input parameter are derived from the collector- and base-dopant profiles. The model parameter extraction can be used to investigate the device structure. Device's

structures are usually obtained with reverse-biased differential C-V method or SIMS. Accuracy of the C-V method is generally influenced by the junction abruptness and Debye length [32]. The applicable range of the C-V method is limited by junction breakdown voltages. SIMS is reasonably accurate. However, SIMS generally demands very large test structures; the results thus represent predominant 1D diffused dopant profile which may not be applicable directly to practical-size devices. For instance, because of the 2-D diffusion effects occurred in manufacturing processes, the emitter-base junction depth of the BJT's from the same process can be significant deviated for different emitter feature sizes, thereby influencing the peak base doping at the base-emitter junction. This effect can be also observed by comparing the electrical characteristics of devices with different emitter feature sizes. As shown in Fig. 4-1 and Fig. 4-2, the high-injection roll-off of the current gain β and unity-gain frequency f_T occur at higher current density for narrower emitter devices, because they have shallower emitter-base junctions due to the 2D diffusion effects, and thus higher peak base doping. This is also another reason why we need to calibrate the QBBJT's model parameters from electrical measurements for better precision, although it is also possible to evaluate them from process information.

We have found that the profile-related QBBJT parameters we extracted correlate well with SIMS measurements. This suggests that when MMPSICE is used in this extraction mode, it provides an inexpensive way to infer profiles from nondestructive electrical measurements. Such an approach could be useful for statistical process control, or for investigating an unfamiliar transistor's fabrication process.

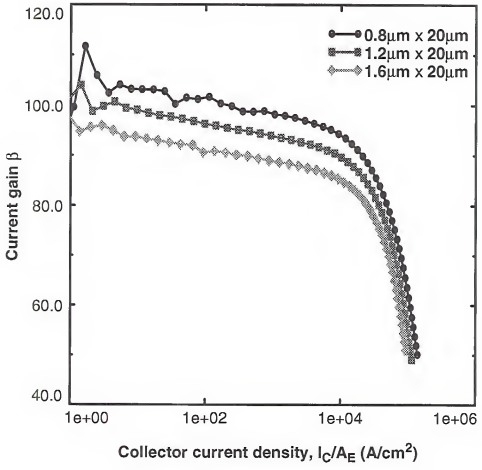


Figure 4-1 Current gain β vs. collector current density J_c for different emitter sizes.

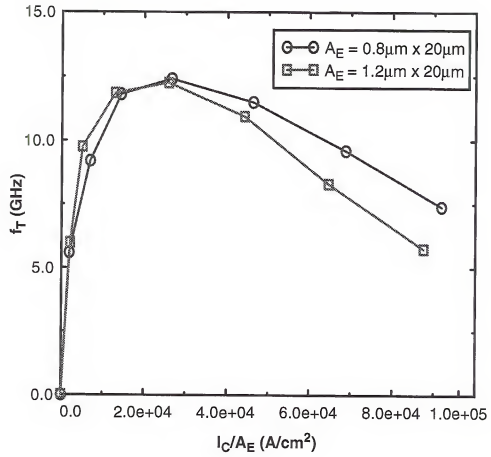


Figure 4-2 Unity-current-gain frequency f_T vs. collector current density J_c for different emitter sizes.

The QBBJT model has 4 device line input parameters and 41 model card parameters. Their physical meanings and typical values are listed in Table 4-1 and Table 4-2. Basically, the QBBJT input parameters can be grouped into the following categories:

1. geometric layout parameters,
2. material dependent carrier transport parameters,
3. Gummel-Poon-like parameters,
4. collector doping profile parameters,
5. base doping profile parameters,
6. charge partition factors.

Table 4-1 Device line input parameters

Name	Description	Units	Default	Typical
AE	Effective emitter area	m^2	1.0e-11	1.0e-11
WE	Effective emitter width	m	1.0e-6	1.0e-6
AC	Effective collector area	m^2	AE	--
ABL	Buried layer area	m^2	AC	--

Table 4-2 Model card parameters

Name	Description	Units	Default	Typical
UNEPI	Electron mobility in epi collector	$cm^2 \bullet V^{-1} \bullet s^{-1}$	1.0e3	1.0e3
NEPI	Epi doping density	cm^{-3}	1.0e16	1.0e16
WEPI	Epi collector width	m	5.0e7	5.0e7
WBM	Metallugical base width	m	2.0e7	2.0e7
NAO	Extrapolated peak base doping	cm^{-3}	1.0e18	1.0e18
ETA	Base doping gradient	-	3.0	3.0
AI	Pre-exponential coefficient for impact ionization rate	cm^{-1}	0.0	7.03e5
BI	Exponential coefficient for impact ionization rate	$V \bullet cm^{-1}$	0.0	1.23e6
VS	Electron saturated drift velocity	$cm \bullet s^{-1}$	1.0e7	1.0e7
JEO	Emitter saturation current density	$A \bullet m^{-2}$	1.0e-8	1.0e-8
JSEO	B-E SCR saturation current density	$A \bullet m^{-2}$	1.0e-4	1.0e-4
NEB	B-E SCR emission coefficient	-	2.0	2.0
WSEO	Zero-bias B-E SCR width	m	5.0e-8	5.0e-8
PE	B-E junction potential barrier	V	1.0	1.0
ME	B-E junction exponential factor	-	0.4	0.4
PC	B-C junction potential barrier	V	0.8	0.8
DNB	Electron diffusivity in base	cm^2	10.0	10.0
PS	C-S junction barrier	V	0.6	0.6
MS	C-S junction exponential factor	-	0.4	0.4
CJS	Zero-bias C-S junction capacitance	$F \bullet cm^{-2}$	0.0	1.0e-4
CRBI	Intrinsic base resistance coefficient	$V \bullet s \bullet cm^{-2}$	2.0e-3	2.0e-4
TB	Carrier lifetime in base	s	1.0e-7	1.0e-7

Table 4-2 Model card parameters (Continued)

TC	Carrier lifetime in collector	s	1.0e-7	1.0e-7
CIF	Forward current coupling coefficient	-	1.0	1.0
CIR	Reverse current coupling coefficient	-	1.0	1.0
RC	Extrinsic collector resistance	W	0.0	50
RB	Extrinsic base resistance	W	0.0	100
RE	Extrinsic emitter resistance	W	0.0	10
FB	Base-charge partition factor	-	1.0	0.5
FC	Collector-charge partition factor	-	1.0	0.5
TE	Emitter hole transit time	s	1.0e-9	1.0e-9
JEOP	Peripheral B-E SCR saturation current density	$A \cdot m^{-1}$	0.0	2.0e-11
NEBP	Peripheral B-E SCR emission coefficient	-	2.0	2.0
UPBASE	Hole mobility in intrinsic base	$cm^2 \cdot V^{-1} \cdot s^{-1}$	230	230
LBE	Width of intrinsic base region	m	0.0	1.5e-6
JCOP	Substrate saturation current density	$A \cdot m^{-1}$	0.0	1.5e-15
DNBH	Electron diffusivity in base for high injection	$cm^2 \cdot s^{-1}$	10.0	10.0
DEGE	Bandgap reduction in base emitter junction	eV	0.0	0.0
DEGC	Bandgap reduction in base collector junction	eV	0.0	0.0
WEM	Emitter junction depth	m	0.0	0.0
DEGA	Average bandgap reduction in emitter	eV	0.0	0.0

Our extraction strategy is described as follows. First, we assume that geometric layout parameters such as the emitter and buried layer dimensions are known. The material-dependent carrier transport parameters, such as majority carrier mobility in the epi-collector, are initially assumed to have their typical values. These values are later refined in the extraction. Then, we extract the Gummel-Poon-like parameters using existing techniques developed for the Gummel-Poon model. Although some of the Gummel-Poon-like parameters are not accurate enough for ac simulation, they are sufficient for the subsequent extraction procedures. With these Gummel-Poon-like parameters, we then extract the collector and base doping profiles through optimization of output characteristics and f_T plot. Finally, we evaluate the charge partition factors and calibrate the extrinsic resistance parameters in accordance with measured small-signal parameters.

To evaluate the charge partition factors and calibrate the extrinsic resistance parameters, we need to extract QBBJT's small-signal equivalent circuit. Generally, there are two different approaches to extract BJT's small-signal circuit. (1) Numerical optimization, which obtains parameters through fitting the equivalent circuit to measured s-parameters. (2) Direct extraction, which obtains parameters step by step through manipulating the small-signal parameters (z, y, or h parameters) derived from the equivalent circuit.

Optimization is a numerical technique used to determine one or more model parameters from measurement. Optimization is especially useful in the fine tuning of multiple parameters to obtain the best model fit to measurements. However, the optimization of multiple parameters is usually not desirable for extracting parameters for a

physical model, because it often results in nonunique or nonphysical parameter sets when given different initial guesses for parameters. There are several methods that prevent optimization from generating nonphysical parameter sets: 1) providing closer initial guesses based on dc parameter extraction; 2) imposing physical constraints; 3) using step-by-step de-embedding techniques with special testing structures [33]; and 4) using advanced optimization algorithms, such as the stimulated diffusion technique [34].

In contrast, direct extractions avoid generating nonunique parameter sets. Therefore, a direct extract methodology is preferred in obtaining a unique set of parameters for a physical model. A direct extraction method is also more computationally efficient than global optimization. Based on the h-parameters and z-parameters derived from BJT's hybrid-T equivalent circuit, some direct extraction methods have been reported [35][36]. However, these methods are not useful to determine QBBJT's small-signal parameters, because they use the over-simplified hybrid-T model. The extraction of small-signal parameters is very sensitive to the topology of the equivalent circuit.

QBBJT's small-signal equivalent circuit is based on the hybrid-π equivalent circuit shown in Fig. 4-3. Recently, a methodology to extract hybrid-π equivalent circuit was reported [37]. The method was also employed to investigate small-signal base impedance [27]. The method is based on the fact that BJT's substrate parasitics are not significant at low frequencies. Using this fact, a BJT's low-frequency z-parameters can be analytically derived from the simplified low-frequency equivalent circuit shown in the dashed box in Fig. 4-3. Through manipulation of the analytic form of BJT's low-frequency z-parameters, small-signal resistances and capacitances such as r_{bb}, r_{ee}, r_{cc} and C_{cc} can be expressed explicitly. However, [37] is not completly direct extraction method, because it does not

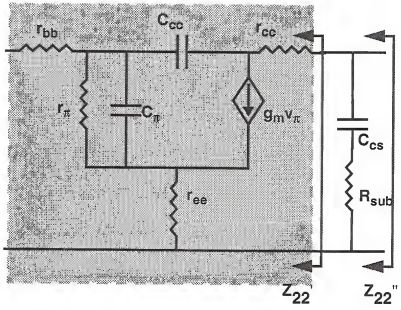


Figure 4-3 The QBBJT's small-signal equivalent circuit

provide a direct way to extract or calculate BJT's base-emitter capacitance C_π. In addition, this method requires a final optimization to obtain the optimal value for the base-collector capacitance partition factor, which accounts for the NQS effects caused by the distributed base impedance.

This chapter presents a complete parameter extraction methodology for the MMSPICE/QBBJT model. With the extraction methodology, QBBJT's parameters can now be evaluated either from process information or from measured data.

The extraction methodology also includes the extraction of BJT's small-signal hybrid-π equivalent circuit from measured s-parameters. With the extracted small-signal equivalent circuit, we can further calibrate the resistance parameters and evaluate the charge partition parameters. The extraction of BJT small-signal equivalent circuit parameters uses no optimization of multiple parameters. Therefore, it ensures the uniqueness of extracted parameters and maintains their physical meanings.

This chapter is organized as follows. Section 4.2 Section discusses the parameters which can be extracted with the techniques developed for the Gummel-Poon model. Section 4.3 discusses the extraction of QBBJT's physical parameters, such as the collector and base doping profiles. Section discusses the extraction of QBBJT's small-signal equivalent circuit. Section 4.5 discusses remaining parameters and concludes this chapter with a summary.

4.2 Extraction of Gummel-Poon-Like Parameters

Some of the QBBJT parameters are analogous to similar parameters in the Gummel-Poon model and can be extracted from a Gummel plot [7]. Saturation collector

current densities J_{CS}, JEO and JSEO are extracted from a linear fitting to the Gummel plots in the medium and low current range. J_{CS} is not a model parameter. J_{CS} will be used to determine DNB during the optimization of the collector doping parameters. Junction capacitance parameters, such as barrier potential and nonideality factor, can be extracted from reverse-bias C-V measurements on large-area devices. These junction capacitances will be used in the optimization of base doping parameters.

Parasitic resistances *RC*, *RB* and *RE* (collector, base and emitter resistances) are required in subsequent extraction and optimizations of collector and base doping profiles. Because QBBJT calculates the voltage drops across the epi-collector and intrinsic base (base region underneath the emitter), thereby including the bias-dependent intrinsic collector resistance and base resistance, *RC* and *RB* represent the extrinsic part of the total collector and base resistances. The intrinsic collector resistance usually dominates the total collector resistance in the forward-active region, but it becomes negligible in hard saturation. The extrinsic collector resistance can therefore be extracted from the slopes of I_C vs. V_{CE} curves in hard saturation (i. e., near $I_C = 0$) [7].

Emitter and base resistances can be extracted simultaneously using Ning and Tang's method [38]. Assuming I_B degrades only due to the V_{BE} drop across the base and emitter resistances, the potential drop can be expressed as

$$V_T \cdot \ln(I_{BO}/I_B)/I_C = RE + (RE + RB)/\beta . \tag{4-1}$$

Therefore, the intercept of the $(V_T \cdot \ln(I_{BO}/I_B))/I_C$ vs. $1/\beta$ curve gives *RE* and the slope gives *(RE+RB)*. A typical base resistance vs. bias plot extracted using this method is shown in Fig. 4-4. The intrinsic base resistance predominates at low injection, but it becomes negligible at high currents because of base majority charge modulation and

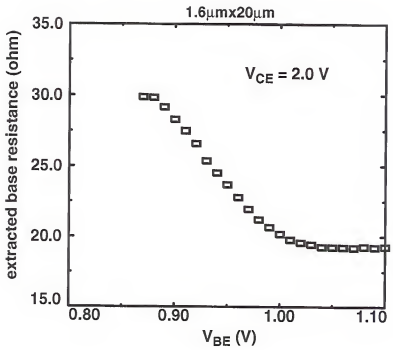


Figure 4-4 Extracted base resistance vs. V_{BE} plot

shrinking of the effective emitter width due to current crowding. Therefore, the extrinsic base resistance can be estimated from the minimum (high-injection) value of the total base resistance.

4.3 Extraction of Doping Profiles

4.3.1 Collector Doping Profile

Collector parameters *WEPI* (epi-layer thickness) and *NEPI* (epi-layer doping) determine the epi-layer resistance, which in turn controls the quasi-saturation I-V characteristics. *NEPI* is directly related to base-collector breakdown voltage as well as the threshold voltage of PMOS devices in n- well BiCMOS technologies [27]. *NEPI* also controls the onset of base pushout and the width of the base-collector SCR, and thus

strongly affects collector transit time. As base widths are progressively scaled down, collector transit time becomes an increasingly significant part of the total transit time and hence strongly affects high-frequency ac characteristics.

Since the electrical performance of a bipolar transistor operating in quasi-saturation is primarily controlled by its collector structure, we can extract collector parameters by optimizing them to fit the QBBJT model prediction to the quasi-saturation I-V characteristics. Unlike most compact BJT models (for example, the extended Gummel-Poon model) model the quasi-saturation operation by assuming that the entire epi-collector is quasi-neutral, QBBJT accounts for all possible SCR formation in the epi-collector. QBBJT solves de Graaff's base transport equation [12] and Kull's collector transport equation [10] subject to moving boundary conditions [17]. Therefore, QBBJT properly simulates current-voltage characteristics accurately in quasi-saturation.

To improve the stability and convergence rate of the optimization, *PC* (B-C junction potential barrier) and R_{EPI} (epi-layer resistance) are used as the target variables for the optimization instead of optimizing *NEPI* and *WEPI* directly. This preserves correlations among the model parameters and helps to eliminate nonphysical parameter combinations. Because the optimization of the collector parameters is fairly insensitive to the base doping gradient, the base grading factor *ETA* can temporarily be assigned an typical value. Assuming the base doping at the base-collector metallurgical junction equals the collector doping, *NEPI*, *WEPI*, and *NAO* (extrapolated peak base doping) can be related to *PC* and R_{EPI} as

$$PC = 2 \cdot \ln(NEPI/n_i) \tag{4-2}$$

$$R_{EPI} = WEPI/(q \cdot UNEPI \cdot A_C \cdot NEPI) \tag{4-3}$$

$$NAO = NEPI \cdot \exp(ETA) \tag{4-4}$$

Calculating *WBM* (metallurgical base width) and *DNB* (average minority diffusivity in base) requires a value for the total base doping, Q_{BO}. This parameter can be extracted from the forward Early voltage if self-heating is eliminated [21]. The Early voltage V_A, defined as $V_A = (1/I_C)\partial I_C/\partial V_{CB}$, can be expressed as [39]

$$|V_A| = \frac{Q_{BO} - Q_{JE} - Q_{JC}}{\varepsilon_S/(W_{SCB} + W_{SCC})}, \tag{4-5}$$

where ε_S is the silicon permittivity, W_{SCB} and W_{SCC} are the base-side and collector-side depletion widths of the base-collector junction, and Q_{JE} and Q_{JC} are the depletion charges of the emitter and collector junctions. Under low-injection conditions (i. e., neglecting background charge modulation induced by collector current) W_{SCC} and W_{SCB} can be calculated using the depletion approximation. Q_{JC} can be calculated as $q \cdot NEPI \cdot W_{SCC}$. With the base-emitter depletion width estimated assuming a one-side abrupt junction, Q_{JE} can be obtained by integrating the base dopant across the base-emitter depletion region. Since V_A is well-defined for every iteration value of Q_{BO}, Q_{BO} can be obtained by optimizing it to fit the V_A-V_{CB} curve.

With Q_{BO} obtained from the previous step, *WBM* and *DNB* can be calculated as

$$WBM = \frac{Q_{BO}}{NAO} \cdot \frac{ETA}{1 - \exp(-ETA)} \tag{4-6}$$

$$DNB = \frac{J_{CS} \cdot N_G}{q \cdot n_{ie}^2} \tag{4-7}$$

where N_G is the Gummel number, defined as $N_G = (Q_{BO} - Q_{JE} - Q_{JC})/q$, and J_{CS} is extracted from the Gummel plot as discussed in Section 2. With all the correlated parameters recalculated for every iteration, the collector parameters can be obtained from optimizing them to fit the model prediction to the measurement data. However, From (4-7), it is clear that variation of the bandgap can critically affect the calculated value of *DNB*. For a contemporary bipolar transistor with a very thin and exponentially doped base, the base doping at the emitter side can be as high as 10^{18} cm^{-3}, causing significant bandgap reduction. Overlooking such bandgap variation would lead to significant errors in the calculation of *DNB* due to the exponential dependence of n_{ie} on the bandgap variation. Extraction of the bandgap variation parameter is possible using temperature-dependent measurement [40]. However, for simplicity we used the empirical model in [41] to estimate the bandgap variation at the emitter edge of the base using the value of *NAO* for every iteration.

To ensure the reasonable values of the bandgap variation parameters, the intermediate values of DNB can be monitored to check the consistency. Based on the estimated base-dopant profile and the minority mobility model in [42], a reasonable range of *DNB* (typically from 6 to 12 cm^2/sec.) can be obtained using Einstein's relation. If the intermediate values of DNB go out of the valid range, the bandgap variation parameters can be fine-tuned in accordance with the calculation of [41]. The flowchart of the optimization procedures of collector parameters is shown in Fig. 4-5.

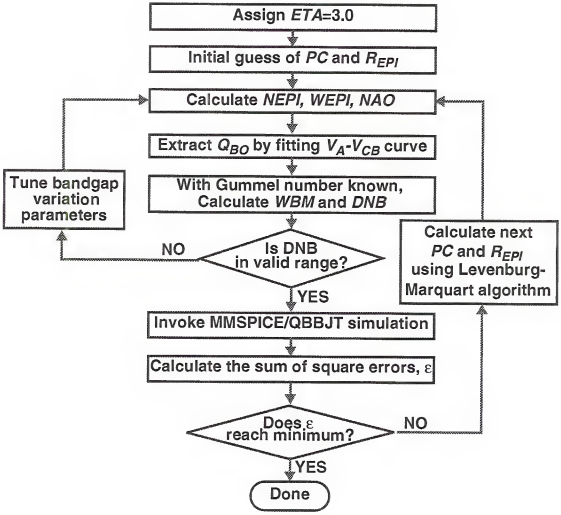


Figure 4-5 Flowchart of collector doping extraction

4.3.2 Base Doping Profile

Although total base doping Q_{BO} can be obtained by fitting the forward Early voltage V_A as a function of V_{BC}, the actual base-dopant grading factor *ETA* cannot be obtained precisely from dc measurements because dc current-voltage characteristics depend only weakly on base-dopant grading. However, the base transit time τ_B depends strongly on the base-dopant grading because of the electric force it causes. QBBJT physically accounts for the dependence of the base transit time on the built-in electrical field, so *ETA* can be extracted by optimizing it to fit f_T data.

The total forward transit time, $\tau_T = (\tau_B + \tau_C + \tau_E)$ can be estimated from the unity current-gain frequency using $\tau_T = 1/(2\pi \cdot f_T)$. For narrow base widths, τ_C and τ_E can become comparable to τ_B and must be carefully accounted for in order to correctly predict f_T.

In its regional charge analysis, QBBJT models τ_C as Q_C/I_C, where $Q_C = W_{SCC} \cdot (qN_{EPI} - I_C/\upsilon_s)$. Using *NEPI* obtained from the collector-profile extraction, W_{SCC} can be calculated based on the depletion approximation, modified to account for the background charge modulation induced by collector current, which allows an accurate estimate of τ_C.

QBBJT uses the empirical parameter *TE* (emitter minority-carrier transit time) to account for τ_E. τ_E can be related to TE as $\tau_E = Q_{EE}/I_C = (TE \cdot I_B)/I_C \approx TE/\beta$, where Q_{EE} is the extra charge storage in the emitter. If the emitter junction is very shallow (< 0.1 μm) and has a metal contact, the extra charge storage in emitter is generally very small and

τ_E can often be neglected. However, if the emitter is thick and uses a polysilicon contact which can accumulate a significant charge, τ_E can be comparable to τ_B and must be calculated independently; if the detailed emitter structure is known, τ_E can be estimated using the formula in [44].

ETA can be extracted by optimizing *ETA* to fit f_T vs. V_{BE} curves using the previously extracted values of *NEPI* and *WEPI*. Since the quasi-neutral base can extend into epi-collector through either ohmic or nonohmic quasi-saturation mechanisms, it is important to select the right bias range for the f_T measurements used in extracting *ETA*. V_{BE} should be restricted to ensure that I_C is smaller than the threshold for base push-out at $I_O = qA\upsilon_S N_{EPI}$ [39], and V_{CB} must be kept high enough to avoid quasi-saturation.

4.3.3 Integration of Doping Profile Extraction Procedures

The optimization procedures described above were implemented using the Optimization Toolbox in MATLAB. A translation program was written in C that converts MATLAB outputs to MMSPICE/QBBJT input parameters, invokes MMSPICE simulations, and returns the simulation results to MATLAB. The MMSPICE code was modified slightly to reduce the number of required file accesses. Also, a provision was added to calculate f_T directly from the regional charges in dc analysis rather than by extrapolation using ac analyses.

The empirical parameters J_{CS}, *JEO*, *JSEO*, and the parasitic *RE*, *RC* and *RB* are extracted from Gummel plots as described in Section II. Then, with a default value assigned to *ETA*, the collector parameters are extracted by fitting to quasi-saturation I-V

curves. Next the emitter transit time τ_E is estimated using the formula in [44], and the base-dopant grading factor *ETA* is extracted from optimization of f_T vs. V_{BE} curves as in Section III. As a final step, the zero-bias base-emitter SCR width *WSEO* is tuned to fit β vs. V_{BE} curves.

4.3.4 Extraction Results and Discussion

The parameter extraction procedures are tested on double-poly self-aligned BJT devices from a Motorola 0.5μm BiCMOS process. Three sets of data are taken. They are 1) Gummel plots, e.g. $\log(I_C)$ and $\log(I_B)$ vs. V_{BE} with V_{CE} as parameter, 2) I_C vs. V_{CE} plots with V_{BE} as parameter, 3) f_T vs. V_{BE} plots. The dc measurements are done with HP4145A semiconductor parameter analyzer. To avoid significant self-heating, the forward Early voltage data, which are used to extract Q_{BO} in optimization of collector structural parameters, are evaluated at lower V_{BE}. In our present case, the forward Early voltages are extracted from I_C vs. V_{CE} curve at V_{BE}=0.8V. Because of their easy-calibration nature, s-parameter measurements, which are used most widely for characterizing small-signal behaviors of devices, are employed to extract f_T data. With the Cascade air co-planner microwave probes, the s-parameter measurements are done using the HP8510A network analyzer set in conjunction with the HP4140 dc power supply source. Since small-signal current gain β can be evaluated from s-parameter data as $\beta = -s_{21}/[(1-s_{11})\cdot(1+s_{22})+s_{12}s_{21}]$ [45], f_T vs. V_{BE} curve can be obtained from extrapolating β vs. frequency plot to unity gain at each V_{BE}.

In spite of taking 2-D effects into consideration[16], e.g. side-wall currents and extrinsic parasitic capacitances, the QBBJT model is virtually a 1-D model. To minimize the influence of 2-D effects on parameters extraction, measurements are done on a considerably large (1.6μm x 20μm) emitter device. Large devices also help to reduce the importance of de-embedding work in s-parameter measurements. However, large devices aggravate emitter current crowding. The current crowding effect causes devices to enter high injection earlier by increasing current densities corresponding to reduced effective emitter area. Therefore, the current crowding makes the β and f_T vs. V_{BE} curves roll-off faster at high current density. Physically accounting for the current crowding as discussed in Chapter 2, the QBBJT model eliminates the drawbacks of parameter extraction on large devices.

The extracted collector parameters and base-dopant grading factor were found to agree closely with corresponding parameters extracted directly from SIMS profiles as shown in Table 4-3. However, the extracted base width was 16% larger than that indicated by SIMS, which led to about a 2:1 error in the peak doping NAO. This discrepancy can be explained by noting that the SIMS measurements were performed on very large area devices. Because of two-dimensional diffusion effects, the emitter junction of small devices is expected to be shallower than that of large devices. Therefore, the base of the small device should be wider and contain more dopants than predicted by SIMS.

Table 4-3 Comparison between SIMS and extraction results

	NEPI	WEPI	NAO	ETA/WBM	WBM
SIMS	4e16 - 1e17	0.50μm	1.00e18	20.10 μm^{-1}	0.140 μm
Extraction	7.4e16	0.47μm	2.12e18	20.36 μm^{-1}	0.163μm
Difference%	--	6%	112%	1%	16.4%

The simulations with extracted parameters are shown in Fig. 4-6 through Fig. 4-9. A new version of the QBBJT model which includes self-heating and modified current crowding model is used to simulate device characteristics with the extracted parameters. The thermal impedance is calculated theoretically as in [22] or evaluated as discussed in [23]. Notice that the extracted physical parameters such as base and collector doping profiles are also applicable to the BJT's of the same structure but with different emitter feature sizes. For instance, NAO, WBM, ETA, NEPI, and WEPI extracted from the 1.6μmx20μm device can be used for the 0.8μmx20μm device without repeating the complete extraction procedures. This suggests that the model is truly scalable and thus can reduce the parameter extraction efforts significantly.

One bonus feature of the parameter extraction methodology is that with the parameters calibrated with measurements, the first-oder derivative discontinuities in QBBJT's collector current can be removed. Depending on the existence of base-collector SCR, QBBJT uses two different sets of equations to calculate the collector current. This can causes the first-order derivative discontinuity in QBBJT's collector current, when operation transits between quasi-saturation and forward-active modes, if given

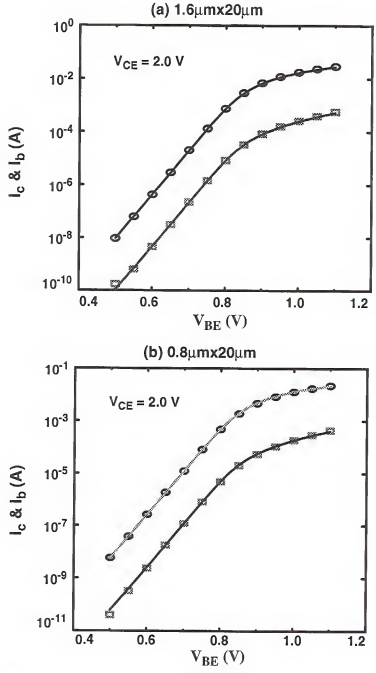


Figure 4-6 Gummel plot, $\log(I_C)$ & $\log(I_B)$ vs. V_{BE}. Symbol lines are the measurement data. Solid lines are simulation results.

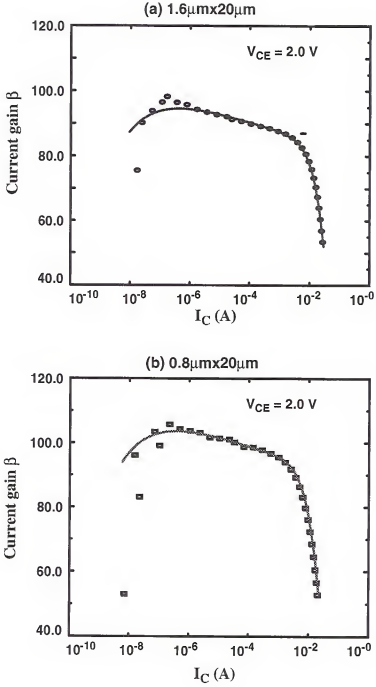


Figure 4-7 Current gain β vs. V_{BE}. Symbol line is the measurement data. Solid line is simulation with current crowding effect. Dashed-line is simulation without current crowding.

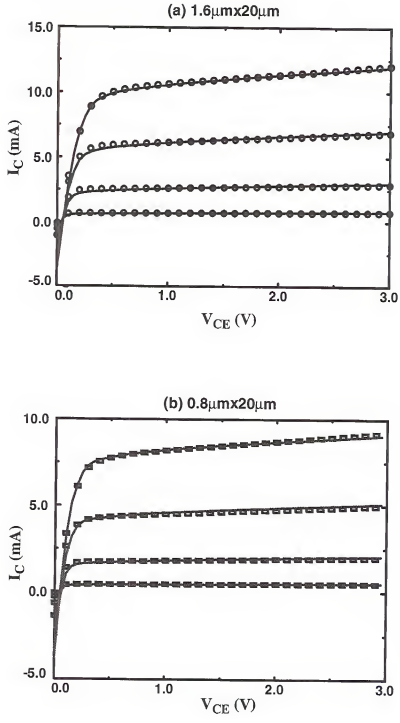


Figure 4-8 I_C vs. V_{CE} plot. Symbol lines are the measurement data. Solid lines are simulation results with self-heating. Dashed-line are simulation results without self-heating.

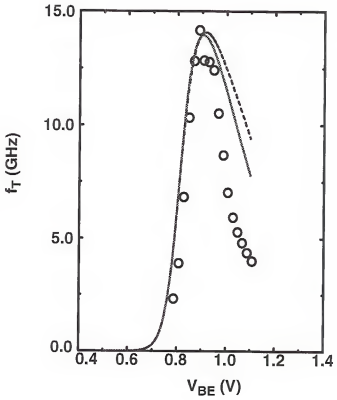


Figure 4-9 f_T vs. V_{BE} plot. Symbol line is the measurement data. Solid line is simulation with current crowding effect. Dashed-line is simulation without current crowding.

inappropriate parameters. Because the collector parameters are obtained from optimization, which in turn conform the model prediction to fit the measurement data, the discontinuity problem can be avoided as shown in Fig. 4-10.

4.4 Extraction of QBBJT's Small-Signal Equivalent Circuit

4.4.1 Methodology

In previous sections, most of QBBJT's parameters were extracted from dc current-voltage characteristics, low-frequency CV measurements and f_T plots. The resulting model parameters give very accurate predictions of BJT's dc characteristics. However, some of these parameters might not be optimal for ac simulations, because isolation of their individual effects from dc measurement is difficult. For instance, the base and emitter resistances can not be accurately determined through dc measurements, because their effects are coupled together at dc. Also, for accurate ac simulation, we still need to know the charge partition factors, which are used in QBBJT to empirically account for the nonquasi-static (NQS) carrier transport in the quasi-neutral base.

This section presents a new methodology to directly extract QBBJT's small-signal equivalent circuit, from which we can evaluate the optimal value of RE, RB, and charge partition factors for accurate ac simulation. The extraction methodology begins with the calculation of the differential current gain and base-emitter resistance from the dc current-voltage characteristics. At medium- to high-injection levels, the base current is dominated by the minority carrier diffusion current in the emitter, given as

$$I_B = I_{BO}\exp(V_{BE}'/V_T), \tag{4-8}$$

where V_{BE}' is internal bias voltage. Therefore, r_π can expressed as

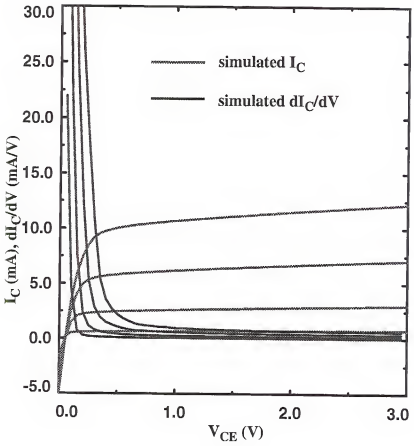


Figure 4-10 Simulated I_C vs. V_{CE} and its first derivative.

$$r_\pi = 1/(dI_B/dV_{BE}') = V_T/I_B \tag{4-9}$$

Because the ideality (exponential) factor of collector current is not necessarily equal to unity at high injection, we cannot directly obtain g_m by direct differentiation of the collector current I_C with respect to the base-emitter voltage V_{BE}'. Instead, we should use

$$g_m = \frac{dI_C}{dV_{BE}'} = \frac{dI_C/dI_B}{dV_{BE}'/dI_B} = \frac{\beta_{ac}}{r_\pi} \tag{4-10}$$

where β_{ac} is low-frequency common-emitter current gain, which can be obtained from BJT's dc collector and base currents. In [37], g_m is expressed as

$$g_m = g_{mo}\exp(-j\omega\tau_d) \tag{4-11}$$

where the exponential term is used to account for the phase-shift caused by the nonquasi-static (NQS) minority transport in the quasi-neutral base. Since such a NQS effect has been accounted for in QBBJT with the charge partition technique, we use (4-10) instead of (4-11).

Since contemporary silicon bipolar transistors are usually built on a very lightly doped substrate or epi-layer, the resulting substrate parasitics are insignificant at low frequencies. With these substrate parasitics neglected at low frequency, BJT z-parameters without consideration of probe pad parasitics are expressed as

$$z_{11} = r_{bb} + r_{ee} + Z_\pi/(1 + g_m Z_\pi) \tag{4-12}$$

$$z_{12} = r_{ee} + Z_\pi/(1 + g_m Z_\pi) \tag{4-13}$$

$$z_{21} = r_{ee} + [Z_\pi/(1 + g_m Z_\pi)](1 - g_m/j\omega C_{cc}) \tag{4-14}$$

$$z_{22} = r_{cc} + r_{ee} + (Z_\pi + 1/j\omega C_{cc})/(1 + g_m Z_\pi) \tag{4-15}$$

where $Z_\pi = r_\pi/(1 + j\omega r_\pi C_\pi)$. Subtracting z_{12} from z_{11} yields the base resistance r_{bb} as

$$r_{bb} = Re(z_{11} - z_{12}). \tag{4-16}$$

Subtracting z_{21} from z_{22} yields the base-collector capacitance C_{cc} as

$$C_{cc} = 1/[\omega \cdot Im(z_{22} - z_{21})]. \tag{4-17}$$

To extract emitter resistance r_{ee}, we multiply the numerator and denominator of the second term of (4-13) by g_m and obtain

$$r_{ee} = Re(z_{12} - \alpha_{ac}/g_m) \tag{4-18}$$

where $\alpha_{ac} = \beta_{ac}/(\beta_{ac} + 1)$ is the low-frequency common-base current gain.

To compute the collector resistance r_{cc}, we cannot directly subtract z_{21} from z_{22}, because z_{21} converted from s-parameters is usually unreasonably large at low frequencies [37]. To compute collector resistance r_{cc}, we subtract z_{12} from z_{22} and obtain

$$r_{cc} = Re\left[z_{22} - z_{12} - \frac{1}{j\omega C_{cc}}\left(\frac{1}{1 + g_m Z_\pi}\right)\right] \tag{4-19}$$

The above equation contains another unknown C_π. To calculate C_π, we use BJT's first dominant pole extracted from the unity-gain frequency f_T as $P_{dominant}=\beta_{ac}/(2\pi f_T)$. From the hybrid-$\pi$ equivalent circuit of Fig. 4-3, BJT's first dominant pole is expressed as

$$P_{dominant} = C_\pi r_\pi + C_{cc}[r_\pi + (1 + g_m r_\pi)(r_{ee} + r_{cc})]. \tag{4-20}$$

Notice that (4-19) and (4-20) are independent and contain only two unknowns r_{cc} and C_π. Therefore, we can solve r_{cc} and C_π from (4-19) and (4-20).

With r_π, g_m, r_{bb}, r_{ee}, r_{cc}, C_{cc}, and C_π calculated with (4-9), (4-10), and (4-16)-(4-20), BJT's low-frequency equivalent circuit (i.e. the circuit in the dashed-box in Fig. 4-3) is directly extracted from low-frequency z-parameters. This equivalent circuit fits closely to measured small-signal parameters at low frequency. However, it does not accurately simulate BJT's small-signal parameters at high frequencies, because this equivalent circuit does not take substrate parasitics into considerations. Therefore, we extract substrate

parasitics from the discrepancy between the simulated and measured data at high frequencies.

The extraction of substrate parasitics begins with the calculation of z22 of the low-frequency equivalent circuit. Then the calculated z22 is compared with the measured z22. From the difference, the substrate parasitics are extracted as

$$z_{22}'' = z_{22}' \| \left(r_{sub} + \frac{1}{j\omega C_{cs}} \right), \tag{4-21}$$

where z22' is the simulated z22 of the low-frequency equivalent circuit and z22" is the measured z22.

In practical situations, only R_{sub} is extracted from this step. We use the C_{cs} measured with low-frequency CV measurements, and extract R_{sub} from fitting (4-21) to the measured z22 from 100MHz to 7.5GHz. Since the optimization contains only one variable, we ensure the uniqueness of R_{sub}, and maintain its physical meaning. The extraction methodology of BJT substrate resistance will also be used in the next chapter to verify the calculation of BJT substrate resistance.

4.4.2 Calibration of QBBJT's Extrinsic Resistance Parameters

The extraction methodology was tested on bipolar transistors of the 0.6μm BiCMOS process of Texas Instruments, Inc. S-parameters are used because they are easier to measure at high frequencies than other small-signal parameters such as z-, y-, or h-parameters. Many calibration and de-embedding techniques are also well-developed for s-parameter measurements. A HP8510C network analyzer system was used to measure BJT s-parameters. These s-parameters were taken from 100 MHz to 7.5 GHz for I_B swept from

1 μA to 1 mA and V_{CE} from 1 V to 3 V. The resulting s-parameters were de-embedded with measurements of open structures. Then, BJT's z-parameters were converted from the deembeded s-parameters.

The base-emitter resistance r_π and transconductance g_m are calculated directly from BJT's dc bias currents using (4-9) and (4-10). Therefore, the r_π and g_m are consistent with the prediction of QBBJT using the parameters extracted from Chapter 4. Other components are extracted with the procedures described previously.

Figure 4-11 shows the extraction of r_{ee}. QBBJT models the emitter resistance with a constant parameter RE. RE can be extracted from Ning-Tang's method or the floating collector technique as discussed previously. However, the accuracy of RE extracted with these dc methods is usually limited, because the effects of emitter resistance and base resistance are coupled in dc measurements. Here, we provide an alternative way to determine RE from ac measurements. Notice from Fig. 4-11 that the ac extracted RE reaches a constant value when I_B is greater than a threshold value. Therefore, we can extract RE from the dashed line in Fig. 4-11.

Figure 4-12 illustrates the extraction of r_{cc}. QBBJT calculates the intrinsic collector resistance internally and leaves the extrinsic collector resistance RC as an input parameter. RC can be extracted from the output characteristics of a BJT operating at saturation region [7]. However, the extrinsic collector resistance RC extracted using this method is usually dependent on I_B, because the intrinsic collector resistance is not totally negligible even if the BJT operates at saturation region. Here, we provide an alternative way to extract RC for a BJT operating at forward active region. Notice from Fig. 4-12 that the total collector resistance extracted from BJT's high-frequency s-parameters is almost a

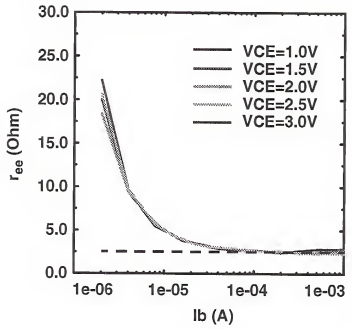


Figure 4-11 The extracted small-signal emitter resistance r_{ee} for different bias conditions.

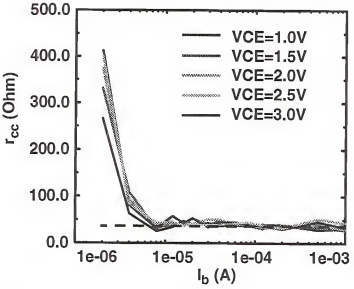


Figure 4-12 The extracted small-signal collector resistance r_{cc} for different bias conditions.

constant when I_B is greater than a threshold value. When a BJT is operating at forward-active region, its intrinsic collector resistance can be approximated as the epi-collector resistance. Therefore, RC can be extracted as

$$RC = RC_{(total)} - R_{EPI}, \tag{4-22}$$

where $R_{EPI} = WEPI/(q \cdot UNEPI \cdot A_C \cdot NEPI)$.

Figure 4-13 illustrates the extraction of total base resistance r_{bb}. Notice that r_{bb} decreases with BJT's base current I_B, and eventually becomes a constant at very high current. This phenomenon is primarily due to the majority charge modulation and dc current crowding. Because the effective base width decreases with base current and the intrinsic resistance eventually becomes zero, the total base resistance approaches the value of the extrinsic base resistance at high current. Therefore, we can extract the extrinsic base resistance RB when the ac extracted total base resistance approaches a constant at high current.

QBBJT calculates the intrinsic base resistance in two different ways. If the current crowding analysis is on, the intrinsic base resistance is internally accounted for in the current crowding analysis as discussed in Chapter 3. If the current crowding analysis is off, the intrinsic base resistance is calculated as

$$RB_{(intrinsic)} = \frac{CRBI \cdot WE \cdot WE}{Q_{BB}}, \tag{4-23}$$

where WE is the emitter width, Q_{BB} is the total base majority charge, and CRBI is an input parameter. In this case, the intrinsic base resistance is a constant for all bias conditions. CRBI can be theoretically calculated as

$$CRBI = \frac{1}{n \cdot UPBASE}, \tag{4-24}$$

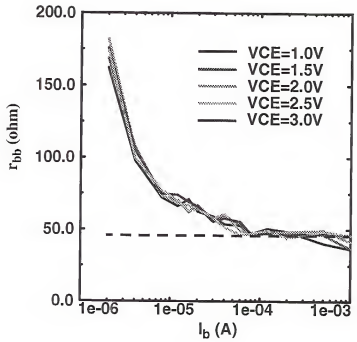


Figure 4-13 The extracted small-signal base resistance r_{bb} for different bias conditions.

where UPBASE is the majority mobility in the base and n is equal to 3 for single base contact configurations and 12 for double base contact configurations. With Q_{BB} calculated from the total base doping for low and medium injection conditions, CRBI can also be tuned to fit the ac extracted total base resistance.

4.4.3 Evaluation of QBBJT's Charge Partition Factors

Figure 4-14 shows the extraction of the base-emitter and base-collector capacitances C_π and C_{cc}. C_{cc} is roughly a constant at low currents, but becomes proportional to C_π for base currents greater than 100μA. This phenomenon is caused by nonquasi-static (NQS) minority transport in the quasi-neutral base. The derivation of the equivalent circuit of Fig. 4-3 is mainly based on the quasi-static (QS) approximation.

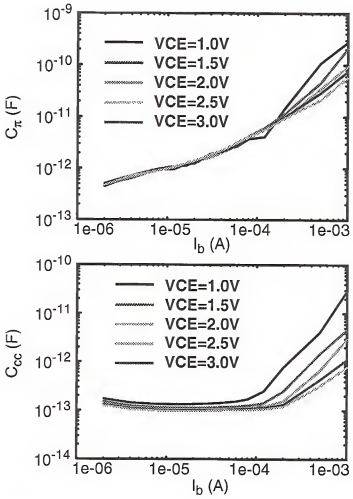


Figure 4-14 The extracted small-signal base-emitter and base-collector capacitances C_{π} and C_{cc} for different bias conditions.

Therefore, the resulting components of the equivalent circuit will be influenced by NQS effects, when the QS equivalent circuit is used to fit the high-frequency measurement. In the present case, the base-collector capacitance C_{cc} is actually comprised of two terms. The first term is the physical base-collector capacitance which can be derived from the depletion charge storage at the base-collector junction. The second term is the NQS term, which is modeled in QBBJT as part of the minority charge storage in the quasi-neutral base. At low current, because the minority charge storage in the quasi-neutral base is very small, the base-collector capacitance is primarily determined by its physical term. However, at high current, the NQS term dominates the value of the extracted base-collector capacitance, because the minority charge storage in the quasi-neutral base becomes much larger than that at low current. Therefore, the extracted base-collector capacitance increases with the base current, when the base current is greater than a threshold value.

QBBJT uses the base charge partition factor FB to account for the NQS effects in the quasi-neutral base. At high current, $C_{cc}/(C_{\pi}+C_{cc})$ is approximately equal to FB. Therefore, we can estimate the base charge partition factor FB as the dashed line shown in Fig. 4-15.

QBBJT uses the collector charge partition factor FC to account for the NQS effects in the extended quasi-neutral base, when base pushout occurs at very high current. Base pushout rarely occur for contemporary high-speed BJTs, since their collectors are usually doped heavily. Also, base pushout is usually not a desired BJT operation. However, if base pushout does occur, FC can be extracted by optimizing the model prediction to the extracted $C_{cc}/(C_{\pi}+C_{cc})$ at very high current.

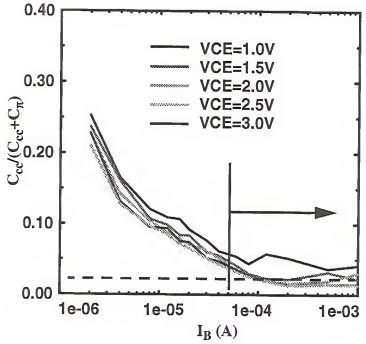


Figure 4-15 Extract the charge partition factor by comparing the C_{cc} and $(C_{cc}+C_R)$ extracted at high current. The dashed line approximates the charge partition factor FB used in QBBJT to account for the NQS carrier transport in quasi-neutral base.

Figure 4-16 shows simulated s-parameters and s-parameters measured from 100 MHz to 7.5 GHz. Fig. 4-17 shows the y-parameter fitting of extracted equivalent circuit. Fig. 4-18 shows simulated unity-gain frequency f_T and f_T computed from measured s-parameter data. We can observe that simulations using the equivalent circuit extracted with the extraction methodology compared very well with measurements.

4.5 Conclusion

The results of this chapter remove a key limitation of the MMSPICE QBBJT model. QBBJT was developed to allow prediction of device/circuit performance from

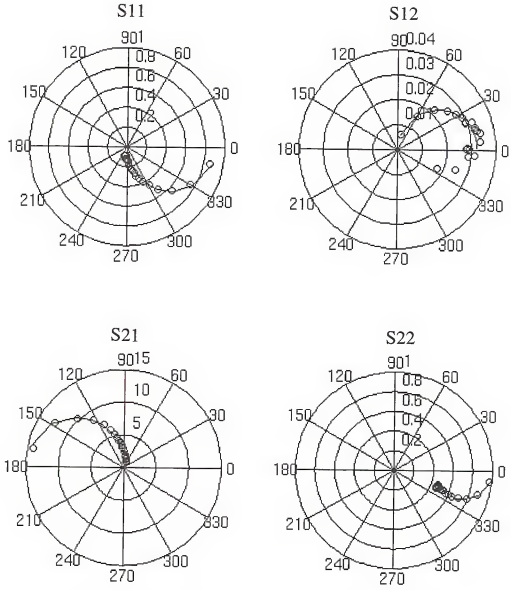


Figure 4-16 The simulated s-parameters compared with the measured s-parameters. Symbol lines represent measurement. Solid lines represent simulation.

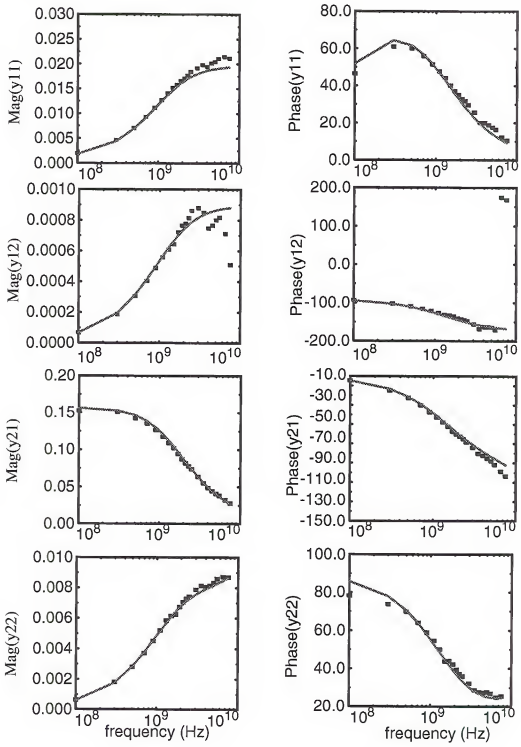


Figure 4-17 The simulated y-parameters compared with the measured y-parameters. Symbol lines represent measurement. Solid lines represent simulation.

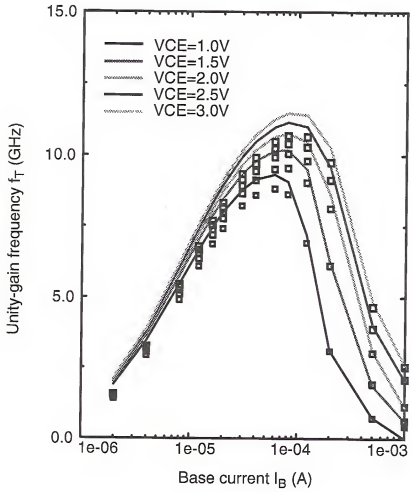


Figure 4-18 The simulated unity-gain frequency f_T compared with the f_T computed from the measured s-parameter data. Symbol lines represent measurement. Solid lines represent simulation.

knowledge of the device structure alone. QBBJT was designed for use in a predictive mode for concurrent engineering applications. This dissertation shows how those physical parameters can be calibrated with the electrical measurements for the representative transistors of a given process. With these calibrated parameters, QBBJT can now be used either in a predictive mode for preliminary circuit design or in a precision mode for final circuit verification.

The methodology begins with the extraction of the Gummel-Poon-like parameters. With these Gummel-Poon-like parameters extracted using the existing techniques, the collector doping profile parameters are extracted from optimizing the output characteristics and the base doping grading factor is extracted from optimizing f_T vs. V_{BE} plots. By demonstrating that the extracted doping profile parameters are highly correlated to the SIMS profiles, this dissertation suggests that the model can be used in a "parameter-extraction" mode where the doping profile parameters are inferred by finding the best fit to electrical measurements.

To further calibrate the extrinsic resistance parameters for accurate ac simulation, a new methodology to extract QBBJT's small-signal equivalent circuit is developed. The method directly extracts BJT's hybrid-π equivalent circuit with using any optimization of multiple parameters. Therefore, this method is more computationally efficient than any global optimization techniques and ensures the uniqueness of the extracted small-signal equivalent circuit. With the extraction of QBBJT's small-signal equivalent circuit, we present a method to extract BJT's charge partition factors for the first time.

Up to now, we have developed a systematic methodology to extract all the QBBJT parameters listed in Table 4-1 and Table 4-2, except AI, BI, CIF, CIR, JCOP, JEOP, NEBP,

and WEM. AI and BI are the pre-exponential and exponential coefficients of the BJT impact ionization rate. AI and BI can be assigned with their default values or extracted from the characteristics of a BJT operating at weak impact ionization region. CIF and CIR are the factors that fine-tune the coupling of the high-injection and low injection components in the forward and reverse collector current. CIF and CIR are usually not sensitive parameters to QBBJT simulation. Therefore, CIF and CIR can also be assigned with their default values. JCOP is the substrate saturation current density, which can be extracted from measuring the forward-biased substrate current. QBBJT uses JEOP, NEBP, and WE to simulate the extra base current injected from the vertical sidewall. These three parameters can be disabled in QBBJT simulation by setting JEOP to zero, since JEO extracted from the total base current has included the sidewall injection term.

This extraction procedure is also adaptable for automated extraction. The measurements required are similar to those needed for extracting Gummel-Poon model parameters. The resulting fits to measurements are at least as close as one normally gets using the Gummel-Poon model, and are typically much better in some regions, such as quasi-saturation. However, with the new extraction methods, the QBBJT model has become much more flexible than Gummel-Poon. Whereas the Gummel-Poon parameters can come only from electrical data, accurate QBBJT parameter sets can be found from many combinations of electrical and physical data, exploiting all available data. Also, the physical QBBJT parameters extracted from one transistor geometry apply naturally and directly to other geometries. Furthermore, as we have seen, the extracted physical parameters can be used to infer dopant profiles, which also suggests a variety of new applications.

CHAPTER 5
SPICE MODELING OF BJT SUBSTRATE RESISTANCE

5.1 Introduction

The previous chapter described a direct extraction method to extract BJT small-signal equivalent circuit parameters including substrate resistance. The method generates an excellent model fit to measured small-signal parameters over a very wide frequency range. However, the extraction must be repeated even for identical BJTs if they have different substrate contact configurations. This chapter presents an efficient modeling methodology for BJT substrate resistance, which is applicable to any BJT model, including QBBJT as well as Gummel-Poon. The methodology eases the burden of repetitive parameter extraction and eliminates the modeling uncertainties caused by substrate resistance.

Substrate resistance is inherent to all semiconductor devices. Substrate resistance causes debiasing effects in MOSFETs and influences the high-frequency ac characteristics of bipolar transistors [48]. As bipolar transistor speed performance continues to improve due to scaling, more and more RF monolithic ICs use silicon bipolar transistors as low cost substitutes for compound semiconductor devices. Substrate resistance can be important in such high-frequency applications. Substrate resistance is, however, not internally accounted for in QBBJT and most compact bipolar transistor models. A conventional approach to compensate for this inadequacy is to optimize the collector-to-

substrate capacitance C_{cs} to empirically account for substrate resistance. This approach, however, is accurate only over a limited frequency range, because it is not physical to represent an RC network using only a collector-to-substrate capacitor [48].

Ignoring substrate resistance limits the accuracy of the simulated small-signal parameter s22 or z22 of BJTs. In broadband RF applications, accurate prediction of s22 or z22 with respect to frequency is essential for design matching networks to compensate the overall gain over a wide range of frequencies. Inclusion of BJT substrate resistance is also essential for predicting the undesired power loss through the substrate and the noise figure (NF) of a low noise amplifier (LNA).

Ignoring substrate resistance also causes numerous problems in BJT parameter extraction. Parameters extracted from two identical transistors with different substrate resistances can vary significantly. This causes difficulties in statistical modeling of parameters and uncertainties in BJT circuit simulations. In addition, parameters extracted from one transistor in a test chip are not necessarily accurate enough for other transistors in a circuit, since these transistors can have various substrate resistance values.

Modeling of substrate resistance is complicated by the layout, i.e., the transistor size, the substrate contact size, and the distance between the transistor and its substrate contact. A typical approach to eliminate the uncertainties caused by different substrate resistances is to carefully control or equalize the values of substrate resistance for all the transistors in the circuit and in the test chip. This can be done by minimizing the values of substrate resistance for all the transistors. To minimize BJT substrate resistance, a guard-ring substrate contact can be placed for each bipolar transistor with a minimum contact-to-device spacing. This approach is reliable and eliminates most of the modeling problems

caused by BJT substrate resistance. However, this method is not desirable because it requires extra silicon area. Minimizing BJT substrate resistance also degrades the small-signal voltage gain at high frequency.

In this chapter, we present an efficient approach to include substrate resistance in BJT parameter extraction and circuit simulation. The method is not associated with any layout rule. It is, therefore, consistent with the goal to minimize the final die area. By including substrate resistance, we can also eliminate the uncertainty caused by substrate resistance in BJT parameter extraction and circuit simulation and ease the burden of repetitive parameter extraction.

The methodology requires an algorithm to accurately calculate substrate resistance from layout and process information. Unlike the current flowing through poly or metal interconnects, which is confined in a well-defined region and can be treated as an one-dimensional current, the substrate current is three-dimensional. This 3-D current flow must be considered in the calculation of substrate resistance.

Our approach is to extend Gharpurey's work on substrate noise coupling to the calculation of BJT substrate resistance [50][51]. The demand for integrated solutions stimulates a growth in mixed-signal products. This growth, in turn, stimulates research on substrate noise coupling [49][50][51]. Substrate noise coupling is a severe problem when analog and digital modules are placed on a same chip [49][50]. The primary goal of this substrate noise coupling research is to quantify the digital noise injection onto sensitive analog nodes through the common substrate. As a key to quantify substrate noise coupling, device-to-device substrate resistance can be computed with the finite difference techniques or boundary element methods. However, these methods are usually

computationally inefficient. Recently, Gharpurey developed a quick algorithm to calculate device-to-device substrate resistances. The algorithm employs the discrete Green function and Fast Fourier Transform (FFT) [51]. Simulations show that Gharpurey's method is as accurate as the finite difference technique, but is 2-3 orders faster [52].

This chapter demonstrates how the recent achievement in the characterization of substrate noise coupling can be applied to the calculation of BJT substrate resistance. Our goal is to eliminate the uncertainties resulting from the layout-dependent substrate resistances and enhance accuracy in SPICE BJT simulation.

To confirm the accuracy of Gharpurey's algorithm in the calculation of BJT substrate resistance, the direct extraction method developed in the previous chapter was used to extract BJT substrate resistance from measured s-parameters. Comparison of extracted and the calculated substrate resistances verifies that the calculation algorithm is accurate.

This chapter is organized as following. Gharpurey's algorithm to calculate substrate resistance is introduced in section 6.2. Limitations of using the algorithm to calculate BJT substrate resistance is discussed in section 6.3. An ac verification for the algorithm is discussed in section 6.4. A systematic methodology to include the layout-dependent substrate resistance in BJT circuit simulation is presented in section 6.5. Finally, a brief conclusion is summarized in section 6.6.

5.2 Algorithm to Calculate BJT Substrate Resistance

To apply Gharpurey's algorithm to the calculation of BJT substrate resistance, we consider the cross-section of a vertical double-poly self-aligned npn bipolar transistor with

a single substrate contact as shown in Fig. 5-1. Since the current flowing along the interface between n-doped region and p-substrate is negligible, the interface (i.e., boundary A) can be assumed to be equipotential. Also, since the region beneath the substrate contact (i.e., the region enclosed by boundary B and substrate contact) is usually highly doped, boundary B can also be assumed to be equipotential. Therefore, we can treat the BJT n-well and the p+ region beneath substrate contact as two metal contacts in our analysis and define the BJT substrate resistance as the resistance between boundary A and boundary B as shown in Fig. 5-1. To calculate the substrate resistance, we supply a unity current into one contact and out of the other, and then calculate the voltage difference between the two contacts. If the substrate doping is uniform, we can assume that the diffusion current in the substrate is negligible. Therefore, substrate current can be expressed as

$$J_{substrate} = q\mu_p pE = \sigma E, \tag{5-1}$$

where μ_p is the hole mobility of the substrate and σ is the conductivity of the substrate. Taking the divergence on both sides of (5-1) yields

$$-\nabla^2\varphi(x, y, z) = \frac{\nabla \bullet J_{substrate}}{\sigma} = \begin{cases} 1/\sigma & \text{at one contact} \\ -1/\sigma & \text{at the other contact} \\ 0 & \text{elsewhere} \end{cases}, \tag{5-2}$$

where φ is the electrostatic potential.

To compute the electrostatic potential resulting from the supplied unity current, we can solve (5-2) using the finite difference method, which discretizes a differential equation into a finite number of difference equations. These difference equations are then linearized with the Newton method, and the resulting coefficient matrix is solved with the sparse-matrix method. The finite difference method is commonly employed in device and process

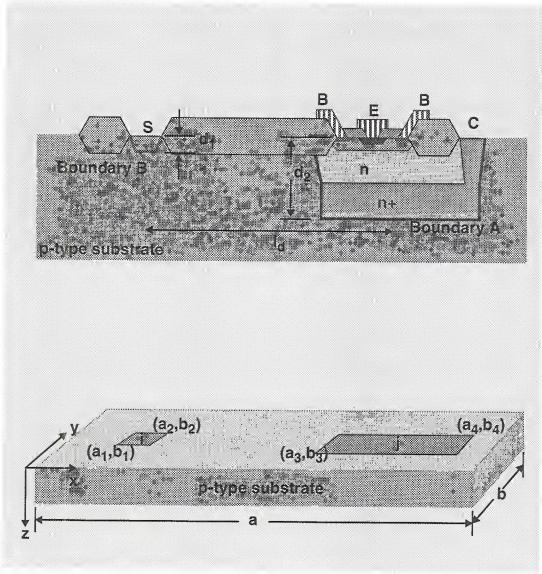


Figure 5-1 Cross-section of a multilayered substrate with two surface contacts.

simulators, such as MEDICI and SUPREME. However, this method is usually very computationally intensive and is thus not feasible when used in the calculation of substrate resistances for a large number of devices.

The Green function technique can also be used to solve (5-2). The Green function technique is often used to obtain the electrostatic potential for a dielectric medium subjected to certain boundary conditions. Green function is defined as the potential at any point (x,y,z) due to a unit point charge locating at (x',y',z'); that is

$$\nabla^2 G(x, y, z, x', y', z') = \begin{cases} 1/\varepsilon & x=x', y=y', z=z' \\ 0 & elsewhere \end{cases}, \tag{5-3}$$

where ε is the permeativity of a dielectric medium. To obtain the electrostatic potential at any point in the space, we can simply sum up the potential contributions from all the charges distributed in the space. That is

$$\varphi(\vec{r}) = \int [\rho(\vec{r}') \cdot G(\vec{r}, \vec{r}')] d\vec{r}'^3 \tag{5-4}$$

The above analysis is developed to compute the electrostatic potential for a dielectric medium. To comprehend how the Green function technique can be applied to computing the electrostatic potential for a resistive medium, we consider two similar situations shown in Fig. 5-2: 1) a purely resistive medium with a unity current flowing into one contact and out of the other; 2) a purely dielectric medium with a positive unity charge on one contact and a negative unity charge on the other. Notice that the boundary conditions of these two cases are mathematically identical except with different proportional constants. Therefore, the electric field or potential of the two cases must be proportional to one another. Therefore, we can also solve the electric potential for a resistive medium with the Green function technique, although it is developed to solve the

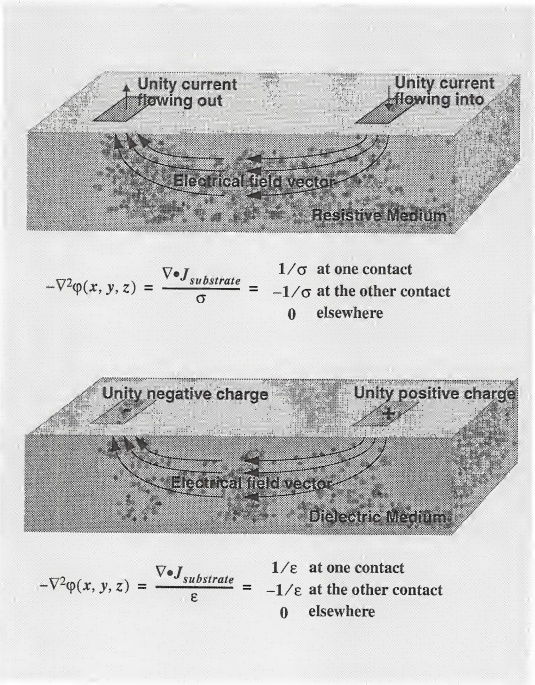


Figure 5-2 The electric field distributions of a purely resistive and a purely dielectric media.

electrostatic potential for a dielectric medium. With the Green function technique, the potential at any point (x,y,z) on the contact j due to a unity current uniformly distributed on the contact i is obtained by integrating the Green function over the whole contact area a_i as

$$\varphi_i(x, y, z) = \frac{1}{a_i}\int_{a_i} G_\sigma(x, y, z, x', y', z')da_i'. \tag{5-5}$$

where G_σ is the Green function for a resistive medium.

Using series expansion and separation of variables, the Green function for a multilayered resistive medium can be expressed as

$$G_\sigma(x, y, x', y')_{z=z'=0} = G_o + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} f_{mn}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{m\pi x'}{a}\right)\cos\left(\frac{n\pi y}{b}\right)\cos\left(\frac{n\pi y'}{b}\right), \tag{5-6}$$

where a and b are the length and width of a rectangular substrate; G_o and f_{mn} are constant coefficients defined by the physical structure of the substrate [51].

The average potential on the contact j due to a unity current uniformly distributed on the contact i can be obtained by integrating (5-5) again over the whole contact area a_j as

$$\varphi_{ij} = \frac{1}{a_i a_j}\int_{a_j}\int_{a_i} G(x, y, z, x', y', z')da_i'da_j'. \tag{5-7}$$

Substituting (5-6) into (5-7) yields

$$\varphi_{ij} = G_o + \left\{\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} f_{mn}\cdot\frac{\left[\sin\left(m\pi\frac{a_2}{a}\right)-\sin\left(m\pi\frac{a_1}{a}\right)\right]\left[\sin\left(m\pi\frac{a_4}{a}\right)-\sin\left(m\pi\frac{a_3}{a}\right)\right]}{(a_2-a_1)(a_4-a_3)}\right.$$

$$\left.\frac{\left[\sin\left(m\pi\frac{b_2}{b}\right)-\sin\left(m\pi\frac{b_1}{b}\right)\right]\left[\sin\left(m\pi\frac{b_4}{b}\right)-\sin\left(m\pi\frac{b_3}{b}\right)\right]}{(b_2-b_1)(b_4-b_3)}\right\} \tag{5-8}$$

Because (5-8) converges slowly, using (5-8) to compute the electrostatic potential is still time-consuming. To make (5-8) converge faster, Gharpurey manipulated (5-8) using the identity of trigonometric function

$$\sin\left(m\pi\frac{a_i}{a}\right)\sin\left(m\pi\frac{a_j}{a}\right)=\left[\cos\left(m\pi\frac{a_i-a_j}{a}\right)+\cos\left(m\pi\frac{a_i+a_j}{a}\right)\right]/2\ , \tag{5-9}$$

and obtained

$$\varphi_{ij}\approx\sum_{m=0}^{P}\sum_{n=0}^{Q}k_{mn}\cos\left[m\pi\left(\frac{p_{1,2}\pm p_{3,4}}{P}\right)\right]\cos\left[n\pi\left(\frac{q_{1,2}\pm q_{3,4}}{Q}\right)\right]\quad , \tag{5-10}$$

where

$$k_{mn}=\frac{a^2b^2}{m^2n^2\pi^4}f_{mn}\,. \tag{5-11}$$

P and Q are integers defining the upper limit of the series summations of (5-10). p_k and q_k represent the integer indexes corresponding to the contact location axes a_k and b_k. p_k and q_k are defined as

$$p_k=integer[(a_k/a)P],\ q_k=integer[(b_k/b)Q]\ . \tag{5-12}$$

Notice that (5-10) can be decomposed into several cosine series summations. Consider the two-dimensional discrete cosine transform (DCT) series K_{pq} of any series k_{mn} as

$$K_{pq}=\sum_{m=0}^{P}\sum_{n=0}^{Q}k_{mn}\cos\left[m\pi\left(\frac{p}{P}\right)\right]\cos\left[n\pi\left(\frac{q}{Q}\right)\right]\,. \tag{5-13}$$

We can express φ_{ij} as a sum of several specific terms in the K_{pq} series. With P and Q chosen to be powers of 2 (i.e., 2^n, n = 1, 2, 3....), we can use the fast fourier transform (FFT) algorithm to facilitate the computation of K_{pq}. Therefore, this method can be much faster than the conventional Green function technique.

5.3 Limitations of the Algorithm

From above discussions, two assumptions are made: 1) the current flowing through a contact is uniformly distributed; and 2) all contacts are planar and located on the top surface of substrate.

The current flowing into a large contact is nonuniformly distributed. Simulations show that the current density around the edge is relatively higher than that in the center. To account for such nonuniformity, a large contact should be partitioned into several small contacts. The substrate resistance of this large contact can then be estimated through evaluating the total conductance of all the small contacts. This technique can also be used to partition a nonrectangular contact. For example, a guard-ring contact can be partitioned as the dashed-lines shown in Fig. 5-3.

As substrate contacts or transistors can penetrate into the substrate with a finite depth as shown in Fig. 5-1, several methods have been developed to account for these vertical extensions [51]. However, all these methods work at the expense of higher computation cost. In practical situations, when the lateral dimension of a contact is much greater than the vertical extension, the current following into the vertical side-wall region can be neglected. Therefore, the substrate resistance can be estimated by accounting for only the current flowing into the bottom area.

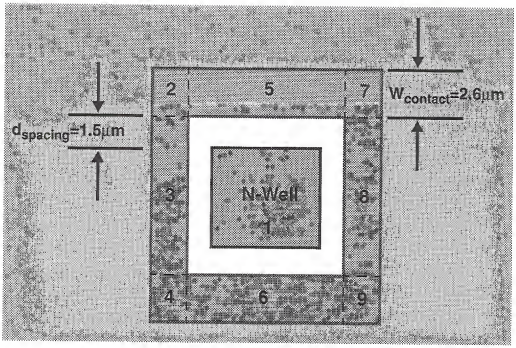


Figure 5-3 Partition of a guard-ring substrate contact

Primary error sources of the Green function technique are the truncation error and the round-off error. The truncation error results when the infinite series summation of (5-10) is truncated with the upper bound P and Q. The round-off error results when (5-12) is used to convert the real numbers a_k and b_k to the integer indices p_k and q_k, which define the contact location in the discrete (integer) domain. To minimize the truncation error, the truncation bound P and Q should be chosen as large as possible. However, this approach makes this algorithm numerically inefficient. In this way, this algorithm loses its advantages over the finite difference method. Experience with this method indicated that (5-10) converges sufficiently when P and Q were equal to or greater than 2^{10}=1024.

The round-off error can be corrected by scaling the calculated resistance by the ratio of the actual contact size to the integer-indexed contact size. The round-off error can

also be completely avoided, if all (a_k, b_k)'s are located at grid points. In this case, we can select substrate size, P, and Q as

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} P \\ Q \end{bmatrix} \cdot d_{grid} \tag{5-14}$$

where d_{grid} is the minimum grid spacing, a and b are the length and width that define a rectangular area to cover two contacts adequately as shown in Fig. 5-1. In this way, the round-off error can be completely eliminated.

5.4 AC Verification of the Algorithm

In [51], Gharpurey had done some dc verifications for the algorithm and applied this algorithm to calculate the equivalent well-to-well (device-to-device) resistances in the characterization of substrate noise coupling. Gharpurey measured the contact-to-contact resistances for various metal contact configurations on a multilayered resistive substrate, and then compared these measured resistances with model predictions. In this section, we present an ac verification for the algorithm applied to the calculation of BJT substrate resistance. Here, we used the direct extraction method developed in the previous chapter. We extracted the substrate resistances for transistors with a wide variety of emitter feature sizes, and then compared these extracted resistances with model prediction.

Test bipolar transistors were manufactured with TI (Texas Instruments, Inc.) 0.6 μm BiCMOS processes. All the test bipolar transistors have a 2.6 μm wide guard-ring substrate contact placed around the BJT n-well with a spacing of 1.5 μm as shown in Fig. 5-3. The detailed layout information is listed in Table 5-1. With the process-specified substrate resistivity (15 ohm-cm), substrate resistances were calculated for transistors of

various emitter feature sizes. S-parameters were measured from 100 MHz to 7.5 GHz and deembedded with open structure.

Fig. 5-4 shows the calculated substrate resistances and the substrate resistances extracted using the method described in the previous chapter. The error bars in Fig. 5-4 represent the standard deviation of the extracted substrate resistances from measurements at different bias conditions. Notice that the extracted substrate resistance values compare well with the calculated substrate resistances for transistors of various emitter feature sizes. This confirms the accuracy of the algorithm used in the calculation of BJT substrate resistance.

Table 5-1 Layout information of the test devices

Test Device W_E x L_E µm x µm	# of parallel devices in test array structure	# of fingers	length of N-well (µm)	width of N-well (µm)
0.6x1.2	54	1	17.8	20.8
0.6x1.5	45	1	19.8	21.7
0.6x3.0	24	1	19.8	23.2
0.6x6.0	12	1	19.8	26.2
0.6x12	6	2	22.8	26.2
0.6x23.9	3	3	25.6	27.5
0.6x36	2	4	28.6	28.8
0.6x45	2	5	31.4	29.2
0.6x72	1	5	31.4	34.0
0.6x96	1	5	31.4	38.8
0.6x192	1	10	46.6	38.8

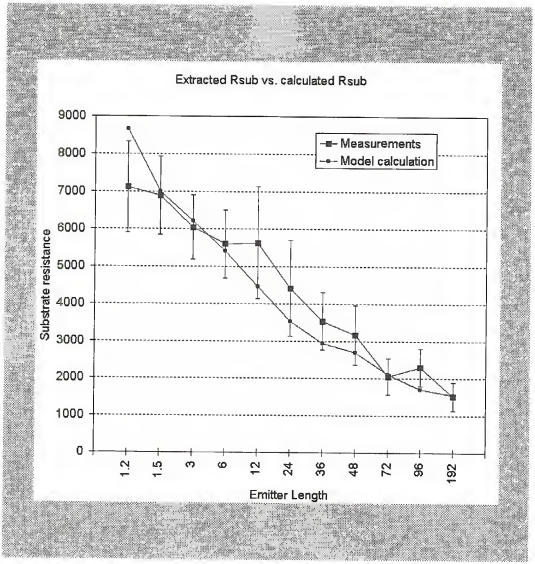


Figure 5-4 Comparison between extracted substrate resistance and calculated resistances for BJTs with a wide variety of emitter feature sizes.

5.5 Including BJT Substrate Resistance in Parameter Extraction

We implemented the methodology in the automated parameter extraction tool of Texas Instruments, Inc., MACH. MACH extraction is based on the conventional Gummel-Poon model. First, MACH extracts dc BJT parameters from Gummel plots and output characteristics. At this stage, MACH also extracts the capacitance parameters from small-signal C-V measurements. Secondly, MACH extracts the forward base transit time by fitting the unity-gain frequency f_T plot. Finally, MACH optimizes RB, RBM, RC, RE, XCJC, and CJS to get the best fit to the measured s-parameters.

Since substrate resistances are not explicitly modeled in MACH, the CJS values obtained from MACH's optimizations can vary even for identical BJTs with different substrate resistances. This causes numerous difficulties in the statistical modeling of CJS. For instance, CJS values obtained from MACH's optimization are often unreasonably small compared to their physical value, because the values are empirically lowered in optimization to account for the substrate resistance.

To include substrate resistance in MACH parameter extraction, we added into MACH a new variable RSUB. To test the methodology, we compared the extraction result of MACH for two different cases: 1) old methodology: set the substrate resistance to zero, and optimize CJS along with RB, RBM, RC, RE, and XCJC; 2) new methodology: use the physical CJS obtained from low-frequency CV measurements and the calculated substrate resistance, and optimize only RB, RBM, RC, RE, and XCJC. The test was performed on the bipolar transistors manufactured using a TI 0.6 μm RF BiCMOS process.

Figure 5-5 shows the differences between extractions using the two different methodologies. Besides the optimal value of RB, RBM, RC, RE, and XCJC, we also

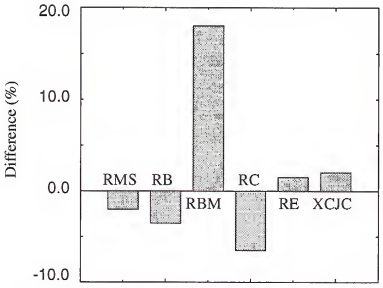


Figure 5-5 Comparison between extractions using new methodology and old methodology. The difference is normalized by the extracted value of the old methodology.

compared the RMS errors of extractions using these two different methodologies. The RMS error is a measure of how well the model prediction fits to the measured data. The percentage differences shown in Fig. 5-5 are normalized by the extraction using the old methodology. Notice that the new methodology fits model predictions to measured data as well as the old methodology. Also, the extraction sensitivities of parameters are very small, which means the extraction of BJT substrate parasitics can be isolated from the extraction of other parasitics. However, the new methodology uses the physical CJS obtained from low-frequency C-V measurement, thereby eliminating in the statistical analysis of CJS the uncertainty caused by the layout dependence of BJT substrate resistance. In addition, with the calculated substrate resistance, we can use the physical

CJS in the extraction, thereby avoiding the need to extract different CJS even for identical BJTs with different substrate resistances.

Notice from that the extraction sensitivity of RBM is relatively large compared with other parameters. This is because RBM is not a sensitive parameter. Very small variation in RB causes relatively large variation in the extracted value of RBM.

5.6 Conclusions

This chapter presents a systematic methodology to include substrate resistance in BJT parameter extraction and circuit simulation. An efficient algorithm to calculate substrate resistance is reviewed. To confirm the accuracy of the algorithm used in the calculation of BJT substrate resistance, substrate resistances were extracted from measured s-parameters and compared with model predictions. The results suggest that the substrate resistance calculated with the algorithm is accurate. The methodology was tested with the TI in-house parameter tool, MACH. The testing results show that with the calculated substrate resistance, we can use the physical CJS in parameter extraction and circuit simulation, thereby avoiding the need to repeat the extraction process for varying substrate contact positions.

The physical collector-to-substrate capacitance and calculated substrate resistance can also be used in QBBJT simulation. Since both the physical collector-to-substrate capacitance and resistance can be calculated from layout and process information, QBBJT thus gains a predictive way to model substrate parasitics.

The algorithm to calculate substrate resistance is very fast. Therefore, it is practical to include the calculated substrate resistance in routine parameter extraction and circuit

CHAPTER 6
SUMMARY AND SUGGESTED FUTURE WORK

6.1 Summary

This dissertation addresses several critical model enhancements and a complete dc and ac parameter extraction methodology for the MMSPICE/QBBJT model. Significant achievements are summarized as follows.

First, self-heating is included in the MMSPICE/QBBJT model. Self-heating significantly influences bipolar transistors even at modest power dissipation. Physical modeling of self-heating depends on accurate simulation of device's local temperature. By adding a thermal node into QBBJT's admittance matrix, the device local temperature is iteratively calculated in the SPICE modified nodal analysis. The resulting current-voltage characteristics are consistent with the device power dissipation and the thermal network attached to the thermal node, thus including the self-heating effect. The method is also adaptable to dc, ac, and transient thermal analysis.

Second, the dc current crowding model is modified to better model the current-crowding induced high-injection effects. The original QBBJT model accounts for dc current crowding around the emitter periphery using the effective base-emitter voltage. This method significantly underestimates the peak injection level around the emitter periphery, thus overlooking the high injection effect caused by dc current crowding. To avoid the underestimation of the high-injection effects caused by dc current crowding, this

dissertation presents an alternative way using the effective emitter area to model dc current crowding. The method is consistent with the previous dc current crowding analysis, but ensuring a correct high injection analysis.

Third, a compact, physical model is developed to account for NQS ac current crowding. This is the first compact model which accounts for NQS majority carrier transport in the quasi-neutral base and the non-uniformity of base-emitter admittance caused by dc current crowding. AC current crowding is observed in BJTs as a nonlinear phase shift in the base input impedance. Adapting the RC transmission line model, this dissertation presents an equivalent circuit which simulates the nonlinear phase shift in the base input impedance. This equivalent circuit is simple and is implemented in the MMSPICE QBBJT model. This model also accounts for the dc base resistance reduction caused by dc current crowding. With the effective base width calculated from the dc current crowding analysis, the equivalent circuit is extended to empirically account for the base-emitter admittance non-uniformity caused by dc current crowding. The extended model is verified with the numerical solution of a nonlinear distributed RC transmission line and the MIDICI simulation.

Fourth, a complete methodology to extract QBBJT's model parameters is presented. With the physical constants (i.e. mobility and carrier saturation velocity) assigned with default values, parameters are extracted in accordance with their significant impacts to transistor's electrical characteristics. The extraction methodology is based on several local optimizations for different operating regions. Because the methodology optimizes no more than two parameters at a time, the methodology ensures the uniqueness of extracted parameters. The physically extracted parameters can, thus, be employed to

infer BJT doping profiles. By demonstrating that the extracted doping profile is closely correlated to the experimental SIMS profile, we suggest that the extraction methodology can be used as an easy, inexpensive, and nondestructive tool for the investigation of BJT doping profiles. In addition, a direct method to extract BJT hybrid-π equivalent circuit is developed. With the direct methodology, the charge partition factor can be extracted and the extrinsic resistance parameters can be further calibrated for accurate ac simulation. The methodology is suitable for an automated extraction process. With the extraction methodology, QBBJT can now be used in either a predictive mode for concurrent engineering applications, or in a precision mode for accurate circuit simulations.

Fifth, this dissertation presents a systematic methodology to include substrate resistance in BJT parameter extraction and circuit simulation. Substrate resistance is often neglected in compact BJT models because it is layout-dependent. Even in most sophisticated parasitics evaluation programs, substrate resistance is still ignored because its calculation involves computation-intensive three-dimensional simulations. Ignoring substrate resistance often causes many uncertainties in circuit simulation and parameter extraction. This dissertation extends Gharpurey's work on substrate coupling to the calculation of BJT substrate resistance. With the calculated substrate resistance, we can use the physical collector-to-substrate capacitance in BJT parameter extraction, thereby eliminating the uncertainties in the statistical analysis of collector-to-substrate capacitance and avoiding the need to repeat the extraction for even identical BJTs but with different substrate resistances.

6.2 Suggested Future Work

This dissertation demonstrates how the MMSPICE/QBBJT model is used in a predictive mode for concurrent engineering as well as in a precision mode as a normal empirical model for routine circuit design verifications. The resulting advantages make this model uniquely flexible. However, until recently, use of MMSPICE is still rather limited. Although lacking advertisements could be one of the major reasons, there is another even more critical factor that limits popularity of QBBJT. It is that the MMSPICE/QBBJT is not as numerically robust as other empirical models. QBBJT calculates terminal currents with different equations. In this way, QBBJT better accounts for moving boundary conditions and transport physics for different operating regions. However, this may cause discontinuities in the derivatives of QBBJT's terminal currents. These discontinuities are harmful to QBBJT's numerical stability and produce problems in some analog simulations, such as harmonic balance analyses. Empirical models usually do not have this problems. However, as most empirical BJT models are also criticized for their improper formulation of collector current and static charge storage in quasi-saturation regions, QBBJT faces the biggest dilemma in its model formulation: physics? or numerical robustness?

To better understand this dilemma, let us first generate a wish list for a desired compact SPICE BJT model and see how the two currently available models (VBIC and QBBJT) can satisfy the list. The results are shown in Table 6-1. Here, VBIC is the representative for empirical models and QBBJT is the representative for physical models. The formulation of VBIC and QBBJT are based on same equations, i.e. de Graaff's base minority transport equation and Kull's collector major transport equation. However, with

different perspectives, the resulting models and their parameters can be very different, i.e. empirical or predictive. Numerical stability and efficiency are guaranteed for empirical models. Scalability and physical plausibility are inherent to physical models. Because of using a single-piece equation to model the majority carrier transport in the epi-collector, VBIC is numerically stable and efficient, but it fails to properly simulate the collector current for all different operating regions. Because of using different equations for different operating regions, QBBJT is physical, but it fails to be continuously differentiable. The major trade-off is to use a single-piece smooth function for numerical robustness or use different equations for physics.

Table 6-1 The wish list for a SPICE BJT model

WISH LIST	VBIC	QBBJT
The model is continuously differentiable.	Yes	No
The model can physically model majority carrier transport in collector for various bias conditions.	No	Yes
The model can include as many secondary effects as possible.	Yes	Yes
The model parameters are scalable.	No	Yes
The model parameters are evaluated from transistor's layout and process information.	No	Yes

Actually, the reason why QBBJT uses different equations is to account for different SCR formations in the epi-collector. Therefore, we suggest to develop a unified collector majority transport equation, which can account for various SCR formations in the epi-collector and ensure the numerical smoothness of resulting terminal currents. In this way, we retain all the advantages of QBBJT, but eliminate all its numerical problems.

Another way to overcome this modeling problem is to investigate whether all SCR formations need to be considered, because they may be insignificant under some circumstances. For instance, base push-out may not happen for some very high-speed BJTs.

Although QBBJT takes several geometric effects, such as current crowding, into considerations, it is predominantly a 1D model. Since 3D device simulations are still very time consuming, especially in accounting for various layout and extrinsic parasitics, we suggest to further enhance QBBJT to account for various quasi-3D geometric effects. This approach can be much more computation efficient than 3D device simulations, but still useful in layout designs.

In Chapter 5, BJT substrate parasitics is modeled as a lumped resistance. The model is experimentally justified up to 7.5 GHz. However, since the substrate parasitics is actually distributed, we expect that the model should be insufficient at higher frequencies. Therefore, we suggest to develop a more realistic model for BJT's substrate parasitics used for applications at higher frequencies.

APPENDIX
THE SUBSTRATE RESISTANCE CALCULATION PROGRAM SUBTEST

SUBTEST is a program written in C to numerically calculate the resistances between two or many plannar contacts locating on the top surface of substrate. It accepts either inputs written in a batch file or interactive inputs. From the specification of the substrate physical structure, the program first calculates the series k_{mn} of (6-11) and its 2-dimensional DCT series K_{pq} using FFT. The program then calculates the resistance between contacts by summing the specific terms of the K_{pq} series in accordance with the contact location indexes (refer to (6-12)). Since k_{mn} and its DCT series K_{pq} depend only on the substrate physical structure regardless of the contact configurations, the K_{pq} series can be reused to calculated the substrate resistances for different contact configurations on the same substrate. This attribute makes the technique especially efficient in such tasks as optimizing contact configuration to minimize the noise injection to some sensitive node through common substrate or calculating substrate resistances for many devices across a very big region. For example, to calculate substrate resistance for all the devices on a big chip as shown in Fig. A-1, we can first divide this big chip into many small regions and calculate the K_{pq} series for one of the subregions. Then, the DCT series is repeatedly used to calculate the substrate resistances for all the active devices all over the chip area.

A sample input file to calculate the substrate resistance of a 0.6μmx96μm BJT with a guard-ring substrate contact is provided in Fig. A-2. The first three lines describe the physical structure of the substrate. It is 102.4μm long, 102.4μm wide, 500μm thick. Line 2 specifies the number of layers. In deriving (6-6), the bottom boundary of a substrate is assumed to be grounded. Therefore, one more layer representing high-resistive air must be mounted to the bottom of a substrate, if the substrate is not grounded on bottom. Line 3 contains the information of the z-axis of the bottom of each layer. Since the z-axis of the first layer bottom is always assumed to 0, this line contains only (# of layers minus 1) numbers specifying the z-axis of the bottoms of the rest of the layers. Line 4 specifies the resistivity of each layer. Line 5 specifies the output file name.

Line 6 and the following lines specify the contact configuration. Line 6 specifies the numbers of contacts. Line 7-16 specify the x and y axes and the number of internal partitions for each contact. The numbers are arranged as x1, x2, y1, y2, # of partitions in x-direction, # of partitions in y-direction. The last line specifies whether to run another contact configuration. If yes, another contact configuration specifications similar to those contained in line 6-16 must be provided. Note that SUBTEST does not allow any text comments in the input files.

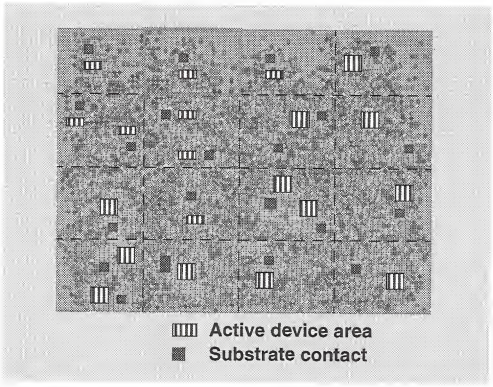


Figure A-1 Divide a big area into many small areas.

```
1:      102.4 102.4 500
2:      2
3:      400
4:      1e8 1e2
5:      sub.out
6:      9
7:      39.6 62.8 35.9 66.5 1 1
8:      35.5 38.1 31.8 34.4 1 1
9:      35.5 38.1 34.4 68.0 1 1
10:     35.5 38.1 68.0 70.6 1 1
11:     38.1 64.3 31.8 34.4 1 1
12:     38.1 64.3 68.0 70.6 1 1
13:     64.3 66.9 31.8 34.4 1 1
15:     64.3 66.9 34.4 68.0 1 1
16:     64.3 66.9 68.0 70.6 1 1
17:     n
```

Figure A-2 Sample input file of SUBTEST

REFERENCES

[1]. J. Warnock, "Silicon bipolar device structures for digital applications: technology trends and future directions," *IEEE Tran. Electron. Dev.*, Vol. ED-42, pp377-389, 1995.

[2]. T. Nakamura, and H. Nishizawa, "Recent progress in bipolar transistor technology," *IEEE Tran. Electron. Dev.*, Vol. ED-42, pp390-398, 1995.

[3]. A.R. Alvarez (Ed.), "BiCMOS technology and applications," *Kluwer Academic Publishers*, Boston, 1989.

[4]. L. Treitinger and M. Miura-Mattausch (Ed.), "Ultra-fast silicon bipolar transistor," *Spring-Verlag*, New York, 1988.

[5]. T. Yamaguchi, S. Uppili, and G. Kawamoto, "Process and device optimization of a 30-GHz f_T submicrometer double poly-si bipolar technology," *IEEE Bipolar Circuits and Technology Meeting*, pp. 136-139, 1993.

[6]. J.J Ebers and J. L. Moll, "Large-signal behavior of junction transistors," *Proc. IRE*, Vol. 42, pp. 1761-1772, 1956.

[7]. I. E. Getreu, "Modeling the bipolar transistor," *Elsevier*, New York, 1978.

[8]. H. K. Gummel and H. C. Poon, "An integral charge control model of bipolar transistors," *The Bell Sys. Tech. J.*, Vol. 49, pp. 827-852, 1970.

[9]. J. G. Fossum, "Modeling issues for advanced bipolar device/circuit simulation," *IEEE Bipolar Circuits and Technology Meeting*, pp. 234-241, 1989.

[10]. G. Kull, L. Nagel, S. Lee, P. Lloyd, J. Prendergast, and H. Dirks, "A unified circuit model for bipolar transistors including quasi-saturation effects," *IEEE Tran. Electron. Dev.*, Vol. ED-32, pp1103-1113, 1985.

[11]. C. McAndrew, J. Seitchit, D. Bowers, M. Dunn, M. Foisy, I. Getreu, M. McSwain, S. Moinian, J. Parker, P. van Wijnen, and L Wagner, "VBIC95: An improved

vertical, IC bipolar transistor model", *IEEE Bipolar Circuits and Technology Meeting*, pp. 170-177, 1995

[12]. H. C. de Graaff and W. J. Kloosterman el. al., "Experience with the new compact MEXTRAM model for bipolar transistors," *IEEE Bipolar Circuits and Technology Meeting*, pp. 246-249, 1989.

[13]. H. C. de Graaff and W. J. Kloosterman, "Modeling of the collector epilayer of a bipolar transistor in the MEXTRAM model," *IEEE Tran. Electron. Dev.*, Vol. ED-42, pp. 274-282, 1995.

[14]. H. Jeong and J.G. Fossum, "A charge-based large-signal bipolar transistor model for device and circuit simulation," *IEEE Tran. Electron. Dev.*, Vol. ED-36, pp124-131, 1989.

[15]. H. C. de Graaff and W. J. Kloosterman, "New formulation of the current and charge relations in bipolar transistors modeling for CAD purposes," *IEEE Tran. Electron. Dev.*, Vol. ED-32, pp. 2415-2419, 1985.

[16]. J. Jin, "Assessment and modeling of nonquasi-static, nonlocal, and multi-dimensional effects in advanced bipolar junction transistors," *Ph.D Dissertation,* University of Florida, Gainesville, 1992.

[17]. G. B. Hong, "Modeling, simulation, and design of SiGe-base heterojunction bipolar transistors for high-performance integrated circuit," *Ph.D Dissertation,* University of Florida, Gainesville, 1994.

[18]. T. Y. Lee, "Extraction of thermal resistance for fully-depleted MOSFETs," *IEEE Int. SOI Conf.*, pp. 78-79, 1995.

[19]. R. M. Fox, S. G. Lee, and D. T. Zweidinger, "The effects of bjt self-heating on circuit behavior," *IEEE J. Solid-State Circuits*, Vol. 28, pp. 678-685, 1993.

[20]. M. Reisch, "Self-heating in BJT circuit parameter extraction," *Solid-State Electron.*, Vol.35, pp. 677-679, 1992.

[21]. G. B. Gao, "Thermal design studies of high-power heterojunction bipolar transistors," *IEEE Tran. Electron. Dev.*, Vol. ED-36, pp. 854-863, 1989.

[22]. J. S. Brodsky, D. T. Zweidinger and Rober M. Fox, "Physics-based multi-pole models for BJT self-heating," *IEEE Bipolar Circuits and Technology Meeting*, pp. 249-252, 1993.

[23]. D. T. Zweidinger and R. M. Fox, "Thermal impedance extraction for bipolar transistors," Accepted for publication by *IEEE Tran. Electron. Dev.*

[24]. R. L. Pritchard, "Two-dimensional current flow in junction transistors at high frequencies," *Proceedings of the IRE*, pp. 1152-1160, 1958.

[25]. H. N. Ghosh, "A distributed model of the junction transistor and its application in the prediction of the emitter-base diode characteristics, base impedance, and pulse response of the device," *IEEE Tran. Electron. Dev.*, Vol. ED-12, pp. 513, 1965.

[26]. J. R. Hauser, "The effects of distributed base potential on emitter-current injection density and effective base resistance for stripe transistor geometries," *IEEE Tran. Electron. Dev.*, Vol. ED-11, pp. 238-242, 1964.

[27]. H. S. Rhee, S. Lee, and B. R. Kim, "DC and AC current crowding effects model analysis in bipolar junction transistors using a news extraction method," *Solid-State Electron.*, Vol.38, pp. 31-35, 1995.

[28]. M. P. J. G. Versleijen, "Distributed high frequency effects in bipolar transistors," *IEEE Bipolar Circuits and Technology Meeting*, pp. 85-88, 1991.

[29]. H. Cho and D. E. Burk, "A simple model for distributed base impedance with ac verification using S-parameter measurements," *IEEE Bipolar Circuits and Technology Meeting*, pp. 106-109, 1990.

[30]. J. G. Fossum, and S. Veeraraghavan, "Partition-charge-base modeling of bipolar transistors for nonquasi-static circuit simulation," *IEEE Electron. Dev. Lett.*, Vol. EDL-7, 1986.

[31]. D. J. Roulston, "Bipolar semiconductor devices," *McGraw-Hill*, pp. 251, New York, 1990.

[32]. W. C. Johnson and P. Panousis, "The influence of debye length on the C-V measurement of doping profiles," *IEEE Tran. Electron. Dev.*, Vol. ED-18, pp. 965, 1971.

[33]. D. Costa, W. U. Liu, and J. S. Harris, "Direct extraction of the AlGaAs/GaAs heterojunction bipolar transistor small-signal equivalent circuit," *IEEE Tran. Electron. Dev.*, Vol. ED-38, pp. 2018-2024, 1991.

[34]. M. Kim, D. Yoon, S. Cha, and J. Jin, "A global optimization of bipolar model parameters using stimulated diffusion," *Proc. IEEE 1994 Int. Conf. on Microelectronic Test Structures*, Vol.7, March 1994.

[35]. D. R. Pehlke and D. Pavlidis, "Evaluation of the factors determining HBT high-frequency performance by direct analysis of S-parameter data," *IEEE Tran. Electron. Dev.*, Vol. ED-39, pp. 2367-2373, 1992.

[36]. S. J. Spiegel, D. Ritter, R. A. Hamm, A. Feygenson, and P. R. Smith, "Extraction of the InP/GaInAs Heterojunction Bipolar Transistor Small-Signal Equivalent Circuit," *IEEE Tran. Electron. Dev.*, Vol. ED-42, pp. 1059-1064, 1995.

[37]. S. Lee, B. R. Ryum, and S. W. Kang, "A new parameter extraction technique for small-signal equivalent circuit of polysilicon emitter bipolar transistors," *IEEE Tran. Electron. Dev.*, Vol. ED-41, pp. 233-238, 1994.

[38]. T. Ning and D. Tang, "Method of determining the emitter and base resistance of bipolar transistors," *IEEE Tran. Electron. Dev.*, Vol. ED-31, pp. 409-412, 1984.

[39]. R. S. Muller and T. I. Kamins, "Device electronics for integrated circuits," 2nd Ed., Wiley, New York, 1986.

[40]. J. W. Slotboom and H. C. de. Graaff, "Measurements of bandgap narrowing in si bipolar transistors," *Solid-State Electron.*, Vol.19, pp. 857-862, 1976.

[41]. D. B. M. Klaassen, J. W. Slotboom and H. C. de Graaff, "Unified apparent bandgap narrowing in n- and p-type silicon," *Solid-State Electron.*, Vol.35, pp. 125-129, 1992.

[42]. D. B. M. Klaassen, "A unified mobility model for device simulation-i. model equations and concentration dependence," *Solid-State Electron.*, Vol.35, pp. 953-959, 1992.

[43]. M. Ugagin and J. G. Fossum, "Significant Time constants defined by high current charge dynamics in advanced silicon-based bipolar transistors," *IEEE Tran. Electron. Dev.*, Vol. ED-41, pp. 1796-1799, 1994.

[44]. K. Suzuki, "Emitter and base transit time of the polycrystalline silicon emitter contact bipolar transistors," *IEEE Tran. Electron. Dev.*, Vol. ED-38, pp. 2512-2518, 1991.

[45]. G. Gonzalez, "Microwave transistor amplifiers - analysis and design," Prentice Hall, Englewood Cliffs, 1984.

[46]. G. A. Armstrong and W. D. French, "A model for the dependence of maximum oscillation frequency on collector to substrate capacitance in bipolar transistors," *Solid-State Electron.*, Vol.38, pp. 1505-1510, 1995.

[47]. A. P. Laser and D. L. Pulfrey, "Reconciliation of methods for estimating fmax for microwave heterojunction transistors," *IEEE Tran. Electron. Dev.*, Vol. ED-38, pp. 1685-1692, 1991.

[48]. M. Pfost, H. Rein, and T. Holzwarth, "Modeling of the substrate effect in high speed s-bipolar ICs," *IEEE Bipolar Circuits and Technology Meeting*, pp. 182-185, 1995.

[49]. N. Verghese, D. Allstot, M. Wolfe, "Verification techniques for substrate coupling and their application to mixed-signal IC design," *IEEE Journal of Solid-State Circuits*, Vol. 31, pp 354-365, 1996.

[50]. R. Gharpurey, and R. Meyer, "Modeling and analysis of substrate coupling in integrated circuits," *IEEE Journal of Solid-State Circuits*, Vol. 31, pp 344-353, 1996.

[51]. R. Gharpurey, *Ph.D. Dissertation*, University of California, Berkeley, 1995.

[52]. R. Gharpurey, and R. Meyer, "Analysis and simulation of substrate coupling in integrated circuit," *International Journal of Circuit theory and Applications*, Vol. 23, 381-394, 1995.

BIOGRAPHICAL SKETCH

Tzung-Yin Lee was born in Kaohsiung, Taiwan in 1969. He received his B.S. degree from the Department of Electronics Engineering at National Chiao Tung University, Hsinchu, Taiwan, in 1991 and M.E. from the Department of Electrical Engineering at the University of Florida, Gainesville, Florida, in 1992.

After obtaining his M.E., Tzung-Yin continued his study in the same department to pursue the degree of Ph.D. Tzung-Yin is currently engaged in the physical modeling for the advanced high-speed bipolar transistors.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert M. Fox, Chairman
Associate Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jerry G. Fossum
Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark E. Law,
Associate Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

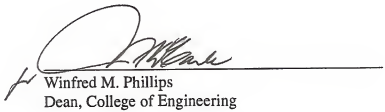
Gijs Bosman
Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mang Tia
Professor of Civil Engineering

This dissertation was submitted to the Graduate Faulty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1997



Winfred M. Phillips
Dean, College of Engineering

Karen A. Holbrook
Dean, Graduate School